School of Economic Sciences

Working Paper Series
WP 2010-20

Keeping Negotiations in the Dark: Environmental Agreements under Incomplete Information

By

Ana Espinola-Arredondo and Felix Munoz-Garcia

June 2012

WASHINGTON STATE UNIVERSITY
World Class. Face to Face.

Keeping Negotiations in the Dark:

Environmental Agreements under Incomplete Information[*]

Ana Espínola-Arredondo[†]
School of Economic Sciences
Washington State University
Pullman, WA 99164

Félix Muñoz-García[‡]
School of Economic Sciences
Washington State University
Pullman, WA 99164

May 30, 2012

Abstract

This paper investigates the role of uncertainty as a tool to support cooperation in international environmental agreements. We consider two layers of uncertainty. Under unilateral uncertainty, treaties become successful with positive probability in the signaling game, even under parameter conditions for which no agreement is reached under complete information. Under bilateral uncertainty, a separating equilibrium emerges where countries participate in the treaty. We then demonstrate under which conditions further layers of uncertainty are welfare improving.

KEYWORDS: Signaling games; Unilateral uncertainty; Bilateral uncertainty.
JEL CLASSIFICATION: C72, D62, Q28.

[*]We would like to especially thank Gene Gruver, Bouwe Dijkstra and Robert Rosenman for their insightful comments and suggestions. We are also grateful to all participants of the the 8th Meeting on Game Theory and Practice Dedicated to Global and International Issues at the University of California, Riverside.

[†]Address: 111C Hulbert Hall, Washington State University, Pullman, WA 99164. E-mail: anaespinola@wsu.edu.

[‡]Address: 103G Hulbert Hall, Washington State University. Pullman, WA 99164-6210. E-mail: fmunoz@wsu.edu. Phone: (509) 335 8402. Fax: (509) 335 1173.

1 Introduction

In the recent 2009 United Nations Climate Change Conference, several observers highlighted the presence of asymmetries in countries' abatement costs, which affects their capacity to comply with the terms of an environmental agreement. In particular, the penetration rate of clean technologies —such as carbon capture and storage, or biodegradable products— was significantly different among countries involved in the negotiations, suggesting a varied ability to fulfill the ambitious reduction in emissions specified in the treaty. This ability is, however, not publicly observable. Since international environmental agreements (IEAs) usually target overall emission levels, requiring the adoption of clean technologies by several industries in a country's economy, the precise dissemination of these technologies along all industries is difficult to observe by outsiders. This dissemination can, however, be more accurately assessed by local governments. At first glance, this information asymmetry could lead one to predict that the negotiation of an environmental agreement will likely be unsuccessful. This paper shows that, in contrast, information asymmetries can actually facilitate the signature of an IEA.

Our paper examines the uncertainty countries face when negotiating IEAs where every country does not perfectly observe other countries' ability to comply with the terms of a treaty that specifies stringent pollutant reductions. In particular, we consider two types of countries: those where the use of clean technologies is widely spread, helping them easily comply with the terms of the IEA (which we refer as "high-technology" countries); and those where these technologies are sparsely available and cannot fully comply with the initial content of the agreement (referred to as "low-technology" countries).[1] Specifically, we study how incomplete information about other countries' widespread use of clean technologies affects countries' participation in the agreement.

We model the signature of the IEA as a signaling game between a leader and a follower, allowing for two layers of uncertainty: one in which only the follower is uninformed about the technological dissemination within the leading country, which we refer as "unilateral uncertainty;" and another where both leader and follower are uninformed about each others' technological dissemination, denoted as "bilateral uncertainty." Unilateral uncertainty explains strategic settings in which the country acting as the follower has a well-known history of fulfilling the content of IEAs due to a widespread use of clean technologies, while the leader's technological dissemination is more difficult to assess.[2] Bilateral uncertainty instead describes cases where, for instance, both leader and follower have recently developed their green industries, and thus a country cannot precisely evaluate to which extent clean technologies are used in the other country's economy. We show that, in the context of unilateral uncertainty, only a pooling equilibrium can be sustained where both types of

[1] Given their low technologies, these countries can introduce an amendment to the treaty that allows them to implement a lower reduction in pollutants. Most international agreements generally allow for such amendments. For instance, the Montreal protocol of 1987 has subsequently been adjusted and amended. Similarly, countries adopted an amendment to Annex B to the Kyoto Protocol in November 2006, and the International Convention for the Regulation of Whaling was also amended in several occasions.

[2] For instance, Brazil's leading role in several IEAs —as recognized by the European Commission (Directorate General for the Environment)— illustrates this information setting, where European countries were uncertain about the degree of dissemination of clean technologies in the Brazilian economy.

leader choose to participate in the agreement and the follower responds by signing the treaty if the probability of facing a high-type leader is sufficiently large. Otherwise, only a semi-separating equilibrium exists where the low-technology leader randomizes between signing and not signing the agreement, and the follower also responds choosing a mixed strategy after observing a signature. Hence, the sheer possibility that the leader has a high technology induces the uninformed follower to sign the IEA with positive probability. Thus, both countries are likely to participate in the treaty, not only when the leader's type is high but also when it is low. This result gives hope for environmental negotiations in practice, since the presence of uncertainty does not necessarily ruin the chances of a successful IEA.

We then investigate the case where all countries are uninformed about each others' types, i.e., bilateral uncertainty. We show that not only a pooling but also a separating equilibrium can be sustained in which only the high-technology leader signs the agreement.[3] Hence, the introduction of an additional layer of uncertainty allows for a finer dissemination of information to the follower (who can perfectly infer the leader's type), but equilibrium outcomes resemble those under complete information, implying that the treaty is only successful if the leader's technology is high. Therefore, the introduction of a single layer of uncertainty allows for agreements to be signed under larger conditions than in complete information, whereas the addition of further layers of uncertainty could hinder the success of the treaty when the follower's technology is likely to be low.

We also provide welfare comparisons of our equilibrium results, showing that the pooling equilibrium yields a larger social welfare than the separating equilibrium when at least one country's technology is high. Since signature patterns in the separating equilibrium coincide with those under complete information, our results imply that settings where countries are uninformed can yield a larger social welfare. Hence, countries' lack of information during the negotiation stage of an environmental agreement can actually be beneficial. However, if the technologies of both countries are low, social welfare under the separating equilibrium exceeds that in any other equilibrium outcome.

Finally, we extend our model to IEAs negotiated between one leader and multiple followers. In particular, we demonstrate that the set of parameter values supporting the pooling equilibrium where all followers participate in the agreement shrinks as more potential signatories are included in the negotiations. Intuitively, free-riding incentives become stronger, making the pooling equilibrium in which all countries sign the agreement sustainable under a more restrictive set of parameter conditions. Nonetheless, "partially cooperative" equilibria now emerge in which a subset of followers participates in the agreement.

Related literature. The literature on IEAs has extensively analyzed negotiations under complete information; see Barrett (1994a, 1994b and 1999) and Cesar (1994). These studies show that more countries decide to participate in international agreements when the difference between the net global benefit in the noncooperative and the full cooperative outcomes is small, i.e., free-riding incentives are small.[4] International negotiations, however, often occur in incomplete information

[3] We also show that these equilibria in pure strategies survive the Cho and Kreps' (1987) Intuitive Criterion.

[4] This literature was extended by models allowing countries to impose "sanctions" on defecting countries, Barrett

contexts, a setting often overlooked by the existing literature.

This paper connects with the literature introducing incomplete information in international negotiations. For instance, Iida (1993) analyzes international agreements using a repeated bargaining game. Specifically, he assumes that a country is uninformed about other countries' status quo, and therefore it cannot perfectly anticipate whether its offers will be accepted in the negotiation. In contrast, we consider that countries are uninformed about each others' technological dissemination, and hence cannot accurately infer whether other signatories will fully comply with the agreement. In addition, we examine both the negotiation and posterior implementation of the IEA, allowing both for unilateral and bilateral uncertainty. Similarly, the literature on international trade has recently examined tariff agreements where countries are privately informed about each others' internal political pressures, Bagwell (2009) and Martin and Vergote (2008), or about the extent to which the import-competing sector is affected by an efficiency shock, Lee (2007).

Furthermore, our conclusions are also related to those of Kreps et al. (1982), who consider the role of informational asymmetries about players' types in the Prisoner's Dilemma game. Specifically, in their model players assign some probability to their opponent playing a conditionally cooperative, tit-for-tat strategy. They show that there is a sequential equilibrium in which players choose to cooperate with positive probability. We similarly demonstrate that the presence of incomplete information about countries' types may lead to cooperation in situations where such equilibrium outcome would not exist among perfectly informed countries. Finally, our results also relate with studies about participation in IEAs, such as Von Stein (2008), who empirically shows that the introduction of flexibility provisions promotes ratification. Our model hence provides a theoretical support for these empirical results, by demonstrating that countries are more willing to participate in international agreements that allow for the introduction of subsequent amendments in the treaty, conditional on countries' technology.

The next section describes the model under incomplete information. Section 3 examines the set of equilibria in the case of unilateral uncertainty, whereas section 4 analyzes equilibrium predictions under bilateral uncertainty and offers welfare comparisons. In section 5 we extend our analysis to N followers, and section 6 concludes.

2 Model

Consider the signature and posterior implementation of the IEA as a two-stage game in which, first, the country leading the negotiations (country 1) announces its participation in the environmental agreement. Once this announcement is made, the following country (country 2) chooses whether to participate in the IEA. Afterwards, countries determine their reduction of emissions. Specifically, the time structure of the game is as follows:

1. Nature selects the leader's technology, which is privately observed by the leader but unob-

(1992 and 1994a) and by models linking the negotiations of transboundary pollution with other issues such as free-trade agreements; see Whalley (1991), Carraro and Siniscalco (2001) and Ederington (2002).

served by the follower. The leader's technology is either high, θ_H, or low, θ_L, with associated probabilities p and $1-p$, respectively.

2. After observing its technology level, the leading country decides whether or not to sign the treaty. In particular, if it signs, the leader accepts an agreement which specifies socially optimal emissions for the high-technology country.[5] If the leader does not sign the agreement, the signature game ends.

3. If the leader signs the treaty, the follower chooses to sign or not sign the agreement, given its posterior beliefs about the leader's type. If the follower responds not participating in the IEA, the signature game ends. In this section, assume that the follower's technology is high.[6]

4. After the signature game ends, the low-technology country introduces amendments to the treaty, thereby revealing its type to both the follower and the international organization that enforces the agreement. If, instead, the leader's type is high, no amendments are introduced, which also conveys its type to other players. If the agreement is signed by both countries, leader and follower fully implement the content of the treaty. When the agreement is not signed, both types of leader independently choose a reduction in pollutants that maximizes second-period utility.

Country i's welfare function is $b(x_i, x_j) - c(x_i, \theta_K)$, where $K = \{H, L\}$ and $j \neq i$, and where x_i denotes country i's abatement level, $i = \{1, 2\}$. In particular, abatement costs are increasing and convex in x_i, and marginal costs are lower for the high-technology country, i.e., $c_{x_i}(x_i, \theta_H) < c_{x_i}(x_i, \theta_L)$. In addition, a country's benefit from abatement, $b(\cdot)$, increases in its own abatement efforts, x_i, but at a decreasing rate, i.e., $b_{x_i}(\cdot) > 0$ and $b_{x_i,x_i}(\cdot) \leq 0$; and the marginal benefit of x_i decreases in x_j, $b_{x_i,x_j}(\cdot)$, thus implying that abatement efforts are strategic substitutes.

Unsuccessful treaty. In the case that the leader does not sign the treaty, the agreement is unsuccessful and countries independently select their abatement efforts in the second-period game. In particular, the Nash equilibrium prescribes a pair of abatement efforts $(x_1(\theta_K, \theta_H), x_2(\theta_K, \theta_H))$ that simultaneously solves the first-order condition $b_{x_1}(x_1, x_2(\theta_K, \theta_H)) = c_{x_1}(x_1, \theta_K)$ for the K-type leader and $b_{x_2}(x_1(\theta_K, \theta_H), x_2) = c_{x_2}(x_2, \theta_H)$ for the high-type follower, yielding an equilibrium welfare of

$$\begin{aligned} V_1(NS; \theta_K) &\equiv b(x_1(\theta_K, \theta_H), x_2(\theta_K, \theta_H)) - c(x_1(\theta_K, \theta_H), \theta_K) \quad \text{and} \\ V_2(NS; \theta_K) &\equiv b(x_1(\theta_K, \theta_H), x_2(\theta_K, \theta_H)) - c(x_2(\theta_K, \theta_H), \theta_H), \end{aligned}$$

[5] Specifically, the international organization leading the negotiations, such as the IPCC, assumes that all countries' technology is high during the negotiation stage, thereby specifying demanding commitment levels. After the agreement comes into force, low-technology countries can introduce amendments to the treaty upon revealing the details of their technological development to the IPCC. High-technology countries, instead, could only conceal their superior technology by destroying it, a case our model does not consider.

[6] This assumption is relaxed in the section on bilateral uncertainty, where the follower has either a high or low technology.

for the leader and follower, respectively. When the leader signs the treaty but the follower responds not participating, the agreement is still unsuccessful, and countries select the same abatement levels $(x_1(\theta_K,\theta_H), x_2(\theta_K,\theta_H))$ as above, thereby yielding equilibrium welfare of $V_1(S,NS;\theta_K) \equiv V_1(NS;\theta_K) - NC(\theta_K,\theta_H)$, where $NC(\theta_K,\theta_H)$ denotes the negotiation cost that the θ_K-leader incurs. Therefore, conditional on the follower not participating in the IEA, the leader prefers not to sign the treaty, i.e., $V_1(NS;\theta_K) > V_1(S,NS;\theta_K)$. In contrast, the follower does not incur negotiation costs, and hence $V_2(S,NS;\theta_K) = V_2(NS;\theta_K)$.[7]

Successful treaty. When the follower responds signing the treaty, the IEA comes into force, and countries must comply with the type-dependent abatement levels specified in the agreement. In particular, the international organization that monitors the treaty maximizes the joint social welfare given the specific type of the signatories. If the leader's type is high, the agreement prescribes socially optimal abatement levels $\left(x_1^{SO}(\theta_H,\theta_H), x_2^{SO}(\theta_H,\theta_H)\right)$, which lie above the efforts that countries independently choose when the treaty is unsuccessful, i.e., $x_i^{SO}(\theta_H,\theta_H) > x_i(\theta_H,\theta_H)$ for all $i = \{1,2\}$.

Leader's payoff. In this setting, the leader's equilibrium welfare is

$$V_1(S,S;\theta_H) \equiv b(x_1^{SO}(\theta_H,\theta_H), x_2^{SO}(\theta_H,\theta_H)) - c(x_1^{SO}(\theta_H,\theta_H),\theta_H) - NC(\theta_H,\theta_H).$$

In addition,

$$\begin{aligned} & b(x_1^{SO}(\theta_H,\theta_H), x_2^{SO}(\theta_H,\theta_H)) - b(x_1(\theta_H,\theta_H), x_2(\theta_H,\theta_H)) \\ > & \left[c(x_1^{SO}(\theta_H,\theta_H),\theta_H) - c(x_1(\theta_H,\theta_H),\theta_H)\right] + NC(\theta_H,\theta_H), \end{aligned}$$

which implies that $V_1(S,S;\theta_H) > V_1(NS;\theta_H)$, and hence the leader prefers to sign the treaty. Intuitively, the improved global environmental quality associated with the IEA offsets the costs from the treaty, which arise from the increase in abatement efforts and the negotiation cost. If, instead, the leader's type is low, the agreement specifies socially optimal abatements $\left(x_1^{SO}(\theta_L,\theta_H), x_2^{SO}(\theta_L,\theta_H)\right)$ which are also higher than the Nash equilibrium abatement efforts for this profile of countries' types, yielding a leader's equilibrium welfare of

$$V_1(S,S;\theta_L) \equiv b(x_1^{SO}(\theta_L,\theta_H), x_2^{SO}(\theta_L,\theta_H)) - c(x_1^{SO}(\theta_L,\theta_H),\theta_L) - NC(\theta_L,\theta_H),$$

which, similarly to the case of high-technology, satisfies

$$\begin{aligned} & b(x_1^{SO}(\theta_L,\theta_H), x_2^{SO}(\theta_L,\theta_H)) - b(x_1(\theta_L,\theta_H), x_2(\theta_L,\theta_H)) \\ > & \left[c(x_1^{SO}(\theta_L,\theta_H),\theta_L) - c(x_1(\theta_L,\theta_H),\theta_L)\right] + NC(\theta_L,\theta_H), \end{aligned}$$

also implying that $V_1(S,S;\theta_L) > V_1(NS;\theta_L)$, indicating that environmental benefits from the treaty offset its associated costs.

[7] The follower's negotiation costs are normalized to zero.

Follower's payoff. Regarding the follower, its equilibrium welfare in this setting is

$$V_2(S,S;\theta_H) \equiv b(x_1^{SO}(\theta_H,\theta_H), x_2^{SO}(\theta_H,\theta_H)) - c(x_2^{SO}(\theta_H,\theta_H),\theta_H)$$

when the leader's type is high, and similarly when the leader's technology is low. Hence, let $BS_2(\theta_K,\theta_H) \equiv V_2(S,S;\theta_K) - V_2(S,NS;\theta_K)$ denote the high-type follower's benefit from joining a treaty with a θ_K-type leader. To make the follower's participation decision interesting, let us assume that, after observing a signature by the leader, the follower chooses to participate when the leader's technology is high, i.e., $BS_2(\theta_H,\theta_H) > 0$, since the benefits from the improved environmental quality offset the increase in abatement costs. In contrast, the follower does not sign the agreement when the leader's technology is low, i.e., $BS_2(\theta_L,\theta_H) < 0$, given that the welfare benefits from joining a treaty with a low-type leader do not compensate the increase in abatement costs. (For illustrative purposes, Appendix 1 provides a parametric example of these benefits.)

3 Unilateral uncertainty

According to the above specification, in a complete information setting the follower chooses to participate in the IEA, after observing a signature of the treaty, when the leader's type is high, but to not sign the agreement when the leader's type is low. This complete information environment hence leads to the standard "pessimistic" result in which IEAs are not signed, unless both leader's and follower's types are high. By introducing incomplete information into this context, we next demonstrate that environmental agreements are signed with strictly positive probability by both countries. Let us first show that there is only one Perfect Bayesian Equilibrium (PBE) involving pure strategies. All proofs are relegated to the appendix.

Lemma 1. *In the IEA signaling game, only a pooling strategy profile can be sustained as a PBE in pure strategies, and it survives the Cho and Kreps' (1987) Intuitive Criterion. Specifically, the leader signs the agreement regardless of its type and the follower responds signing the treaty for any* $p \geq \overline{p}_H$*, where* $\overline{p}_H \equiv \frac{-BS_2(\theta_L,\theta_H)}{BS_2(\theta_H,\theta_H)-BS_2(\theta_L,\theta_H)}$.

First, note that no separating equilibrium can be supported. In particular, the leader cannot choose a type-dependent strategy, in which it signs (does not sign) the agreement when its technology is high (low, respectively). Otherwise, the follower could infer its type by observing its action, and therefore respond as in the complete information setting, where it only joins an agreement signed by the high-technology leader. This would, however, provide the low-technology leader with incentives to sign as well, violating this strategy profile. In contrast, a pooling equilibrium can be supported when the probability of facing a high-type leader is sufficiently large, $p \geq \overline{p}_H$. In this case, the follower responds by signing the agreement, and both types of leader also sign. When the probability of facing a high-technology leader is low, $p < \overline{p}_H$, however, the follower does not participate in the treaty and the former pooling equilibrium cannot be sustained in this signaling game. The following proposition describes countries' mixing strategy profile in this context.

Proposition 1. *A semi-separating strategy profile can be sustained as a PBE of the IEA signaling game, when $p < \overline{p}_H$, in which:*

1. *The leader signs the agreement with probability $p_L = \frac{p}{1-p}\frac{\overline{p}_H}{1-\overline{p}_H}$ when its technology is low, where $p_L \in (0,1)$ if $|BS_2(\theta_H, \theta_H)| > |BS_2(\theta_L, \theta_H)|$, and signs the treaty with probability one when its technology is high, $p_H = 1$; and*

2. *The follower responds by signing the agreement with probability $\widehat{r} \in (0,1)$ where its posterior beliefs are $\mu(H|S) = \overline{p}_H$, where $\widehat{r} \equiv \frac{V_1(NS;\theta_L)-V_1(S,NS;\theta_L)}{V_1(S,S;\theta_L)-V_1(S,NS;\theta_L)}$.*

Thus, the IEA is signed by both types of leader with a strictly positive probability, not only when the leader's technology is high but also when it is low.[8] The introduction of incomplete information about the leader's type, hence, provides a solution to the no-signature result under complete information settings. In particular, if the follower were perfectly informed about the leader's type being low, the treaty would not be signed by either country. However, under unilateral uncertainty, the sheer possibility that the leader might have a high technology induces the follower to sign the IEA with positive probability. As a consequence, both countries are likely to participate in the treaty. Figure 1 depicts the equilibrium predictions of Lemma 1 (pooling PBE) and Proposition 1 (semi-separating PBE). Let us next examine comparative statics of our previous result.

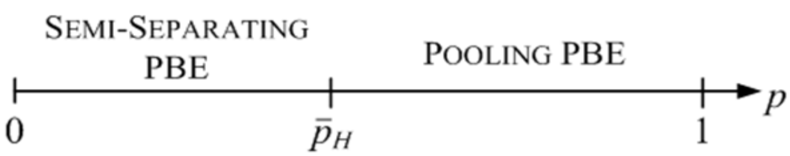

Figure 1. Equilibrium outcomes under unilateral uncercertainty.

Corollary 1. *The leader's probability of signing the agreement when its technology is low, p_L, is: (1) increasing in the probability of the leader being a high type, p; and (2) increasing in the follower's cost of signing a treaty with a low-type leader, $BS_2(\theta_L, \theta_H)$. Furthermore, the follower's probability of signing the treaty, $\widehat{r}$, is: (1) increasing in the leader's negotiation cost, $NC(\theta_L, \theta_H) \equiv V_1(NS;\theta_L) - V_1(S, NS;\theta_L)$; and (2) decreasing in the benefits that the low-type leader obtains when the follower signs the treaty, $V_1(S,S;\theta_L) - V_1(S, NS;\theta_L)$.*

Let us briefly examine the intuition behind the above corollary. First, an increase in the probability of the leader being a high type raises the follower's incentives to sign the treaty, increasing as a result the low-type's probability to participate in the agreement, p_L. Second, an increase in the cost of signing a treaty with a low-type leader reduces the set of beliefs for which the follower

[8] Note that, as described in the proof of Proposition 1, $p_L \in (0,1)$ for all priors p when the follower's benefit from signing the treaty satisfy $|BS_2(\theta_H, \theta_H)| > |BS_2(\theta_L, \theta_H)|$. Otherwise, p_L becomes one, and the semiseparating equilibrium cannot be supported.

responds signing the IEA. Hence, in order to be perceived as a high-technology country, the leader participates in the treaty with a higher probability, p_L. On the other hand, an increase in the leader's negotiation costs makes the signature of the agreement more costly, reducing the likelihood that a signature originates from a low type. As a consequence, the follower is more likely to face a high-technology leader, raising the probability $\widehat{r}$ with which the follower joins the IEA. In contrast, an increase in the benefit that the low-type leader obtains from the follower's signature of the agreement, $V_1(S, S; \theta_L) - V_1(S, NS; \theta_L)$, raises the leader's incentive to participate in the treaty. As a result, the likelihood of facing a low type increases, ultimately reducing the follower's probability of signing the treaty.

4 Bilateral uncertainty

In this section we extend incomplete information to both players. In particular, we consider contexts in which countries negotiate an IEA where every player is privately informed about its own type (dissemination of clean technologies within its jurisdiction), but does not observe the other country's type. This case describes strategic settings where both leader and follower have recently developed their green industries, and every country cannot accurately assess how widespread the use of clean technologies is in the other country's economy. Unlike the unilateral uncertainty case, the follower can now have a high or low technology, with probabilities q and $1 - q$, respectively.

Let us first examine the leader's incentives. Similarly to the previous section, the K-type leader prefers to sign a treaty with a high-type follower, i.e., $BS_1(\theta_K, \theta_H) > 0$, where $BS_1(\theta_K, \theta_J) \equiv V_1(S, S; (\theta_K, \theta_J)) - V_1(NS; (\theta_K, \theta_J))$ and θ_K (θ_J) denotes the leader's (follower's, respectively) type, where $K, J = \{H, L\}$. In addition, when the leader's type is high, the benefits from signing the treaty are positive, regardless of the follower's type, implying that $BS_1(\theta_H, \theta_J) > 0$ for any follower $J = \{H, L\}$, and agreements with a high-technology follower entail larger environmental benefits, $BS_1(\theta_H, \theta_H) > BS_1(\theta_H, \theta_L) > 0$.[9] However, when the leader's technology is low, it prefers to avoid an IEA with a low-type follower, i.e., $BS_1(\theta_L, \theta_H) > 0 > BS_1(\theta_L, \theta_L)$.[10]

Let us now analyze the follower's incentives to sign an IEA. When the follower is a high type, our assumptions from section 2 can be extended to this setting. In particular, the follower's incentives to participate in the agreement with a K-type leader, $BS_2(\theta_K, \theta_H) \equiv V_2(S, S; (\theta_K, \theta_H)) - V_2(S, NS; (\theta_K, \theta_H))$, are positive when facing a high-technology leader, i.e., $BS_2(\theta_H, \theta_H) > 0$, but negative otherwise, $BS_2(\theta_L, \theta_H) < 0$. Intuitively, the environmental benefit from the treaty that the high-technology follower obtains exceeds its associated increase in abatement costs when the leader is also a high-technology country. In contrast, when the leader's type is low, its reduction in emissions is relatively small, thus implying that the benefits the follower obtains do not outweigh its increase in abatement costs. When the follower is a low type, its benefits from participating

[9] This assumption allows for asymmetries in the benefits that leader and follower obtain from the treaty, i.e., the leader's benefits being weakly larger.

[10] This indicates that the increase in abatement and negotiation costs that the low-technology leader must incur outweighs the small environmental benefit of participating in an IEA with a follower whose type is also low, ultimately yielding a negative benefit from signing the agreement.

in a treaty with a high-type leader (where it free-rides the leader's substantial reduction in pollutants) are larger than with a low-type leader, i.e., $BS_2(\theta_H, \theta_L) > BS_2(\theta_L, \theta_L) \geq 0$. The following proposition describes the set of equilibria in pure strategies under bilateral uncertainty.

Proposition 2. *In the signaling game where all countries are uninformed about each others' type, the following equilibria in pure strategies survive the Cho and Kreps' (1987) Intuitive Criterion:*

1. *A separating equilibrium in which the leader signs (does not sign) the agreement when its technology is high (low, respectively), and the follower responds participating (not participating) in the treaty after observing a signature (no signature, respectively), both when its technology is high and low, if and only if* $q < \overline{q}_L$, *where* $\overline{q}_L \equiv \frac{-BS_1(\theta_L,\theta_L)}{BS_1(\theta_L,\theta_H)-BS_1(\theta_L,\theta_L)}$; *and*
2. *A pooling equilibrium in which both types of leader sign the treaty and both types of follower join the agreement if and only if* $p \geq \overline{p}_H$ *and* $q \geq \overline{q}_L$.

The above proposition provides an interesting equilibrium prediction, namely, when all countries are uninformed about each others' types, not only a pooling but also a separating equilibrium can be supported in which information about the leader's type is perfectly transmitted to the follower. This result contrasts with that under unilateral uncertainty where the follower, after observing the leader's signature decision, could not perfectly infer its type in the pooling or semiseparating equilibria of the game, since the leader did not use type-dependent strategies with full probability. Hence, the introduction of an additional layer of uncertainty allows for a finer dissemination of information to the follower. In addition, note that the set of parameter values under which the above separating equilibrium can be sustained depends upon free-riding incentives. In particular, the low-type leader's free-riding incentives are represented by $BS_1(\theta_L, \theta_H)$, since this expression reflects the leader's benefit from signing an agreement with a high-type follower. When such free-riding incentives decrease, the high-type follower is more attracted to participate in the agreement, expanding the set of priors, q, under which the separating equilibrium can be sustained, i.e., shifting $\overline{q}_L$ upwards. (At the end of this section we explore the welfare implications of this equilibrium results and compare them relative to two benchmarks: unilateral uncertainty and complete information.)

The previous proposition describes the set of equilibria under different conditions on the prior probabilities. However, no equilibrium involving pure strategies is predicted in the case that $p < \overline{p}_H$ and $q \geq \overline{q}_L$. The following proposition identifies an equilibrium under these parameter conditions where countries use mixed strategies.

Proposition 3. *In the signaling game where all countries are uninformed about each others' technologies, a semiseparating equilibrium can be supported when* $p < \overline{p}_H$ *and* $q \geq \overline{q}_L$, *where:*

1. *The leader signs with probability* $p_L = \frac{p}{1-p}\frac{\overline{p}_H}{1-\overline{p}_H}$ *when its technology is low, where* $p_L \in (0, 1)$, *and signs the treaty with probability one when its technology is high,* $p_H = 1$; *and*

2. *The follower responds by signing the agreement with probability $r_H(q) \in (0,1)$ when its technology is high, and signing the treaty with probability one when its technology is low, where its posterior beliefs are $\mu_H(H|S) = \overline{p}_H$, and*

$$r_H(q) \equiv \frac{NC(\theta_L, \theta_H)}{BPS_1(\theta_L, \theta_H)} + \frac{1-q}{q}\left(\frac{-BS_1(\theta_L, \theta_L)}{BPS_1(\theta_L, \theta_H)}\right)$$

where $BPS_1(\theta_L, \theta_J) \equiv V_1(S, S; (\theta_L, \theta_J)) - V_1(S, NS; (\theta_L, \theta_J))$ denotes the benefits that the low-technology leader obtains from the posterior signature of the agreement by a J-type follower.

The semiseparating equilibrium predicts that the IEA is signed with positive probabilities. The following figure summarizes the equilibria under different prior probabilities p and q. Specifically, when $q = 1$ the follower has a high technology, i.e., unilateral uncertainty, while $q < 1$ describes our equilibrium results under bilateral uncertainty.

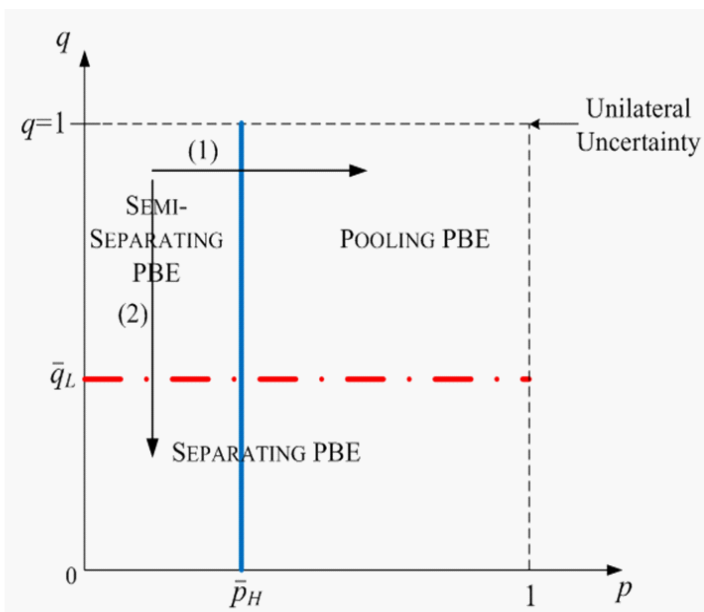

Figure 2. Unilateral versus bilateral uncertainty.

Figure 2 also embodies equilibrium outcomes under complete information. In particular, it describes four possible combinations where: (1) both leader and follower are high type ($p = q = 1$) inducing the leader to sign the treaty and the follower to respond joining;[11] (2) both leader and follower are low type ($p = q = 0$) and the leader does not sign the agreement; (3) the leader is a

[11] This case illustrates the voluntary signature of the Montreal Protocol by several countries. Many specialists argue that the success of this treaty was based on the fact that the technology to reduce CFC gases was already available and widely spread among developed countries (i.e., all countries being high type); see Downs et al. (1996).

high type but the follower is not ($p = 1$ and $q = 0$) inducing the leader to sign the treaty and the follower to respond joining; and (4) the leader is a low type but the follower's type is high ($p = 0$ and $q = 1$) inducing the leader to not sign the treaty.

Let us investigate how the parameters of the model affect the extent of informativeness of the semiseparating equilibrium. A sensible measure is the difference $(p_H - p_L)$. A larger discrepancy in the probabilities of signature by the two types of leader implies that, when a signature is observed, it is more likely to originate from a high-type leader. On one hand, an increase in the probability of the leader being high, p, raises p_L, reducing the degree of informativeness of the semiseparating equilibrium. Graphically, an increase in p moves our equilibrium predictions, from the semiseparating to the pooling equilibrium where both types of leader sign; as illustrated by arrow (1) in the figure. On the other hand, a decrease in the probability of the follower being high-type, q, increases $r_H(q)$. As a result, both types of follower respond signing the agreement, converging to their behavior in the separating equilibrium of the game; as represented by arrow (2). Note that the probability with which both types of leader sign the treaty is independent of q.

4.1 Welfare comparisons

Let us compare the welfare that arises in the equilibrium outcomes of Proposition 2 and 3. In addition, we examine social welfare under different information settings (complete information, unilateral and bilateral uncertainty).

Proposition 4. *Equilibrium welfare satisfies the following ranking:*

1. *When the leader's technology is high, social welfare in the pooling PBE is weakly larger than under all other equilibrium outcomes, $SW_{pooling} = SW_{separ} \geq SW_{semisepar}$, for any follower's type.*

2. *Similarly, when the leader's technology is low but the follower's is high, the pooling PBE yields a larger welfare than any of the other equilibrium outcomes if the environmental benefits from signing the treaty are sufficiently high.*

3. *In contrast, when both countries' technology are low, the pooling PBE produces a lower welfare than all other equilibrium outcomes, $SW_{separ} > SW_{semisepar} > SW_{pooling}$, if the environmental benefits from the agreement are sufficiently low.*

4. *Finally, under complete information, social welfare coincides with SW_{separ}.*

Intuitively, when at least one country's technology is high (as described in points 1 and 2 of Proposition 4), social welfare in the pooling equilibrium is weakly larger than in any other equilibrium outcome. When both countries' technology is high, this welfare ranking is unconditional on the environmental benefits from the treaty, whereas when the leader's type is low this result only holds if such benefits are sufficiently high. On the contrary, when both countries' technology is low, the no signature of the treaty is socially optimal if its associated environmental benefits are low, entailing that the separating equilibrium (in which signature does not occur) yields a larger social welfare than any other equilibrium outcome.

Our results also show that the introduction of incomplete information produces welfare gains only when at least one country's type is high. In particular, when priors are relatively symmetric, i.e., $p \geq \overline{p}_H$ and $q \geq \overline{q}_L$, the pooling equilibrium arises, generating a larger welfare than under complete information. By contrast, when priors are relatively asymmetric, i.e., $p < \overline{p}_H$ and $q \geq \overline{q}_L$, the semiseparating equilibrium emerges, entailing a lower social welfare than in complete information contexts.[12] Finally, when both countries' types are low, the social welfare under complete information is strictly larger than that under the pooling and semiseparating equilibria of the incomplete information game.

5 Extensions

Let us next extend the unilateral signaling game to multiple followers. In particular, consider a time structure in which, first, the informed leader chooses whether to sign the treaty. If the leader signs, N high-type followers individually and sequentially decide whether to participate in the IEA. All uninformed followers infer the same information about the leader's type after observing its signature decision. Specifically, followers can only infer information about the leader's type when the negotiation stage is over. Similarly as in our previous analysis, when the leader does not sign the treaty, the negotiation process fails. In particular, the leader's equilibrium payoff when it does not sign the treaty is $V_1(NS; \theta_K)$ for all $K = \{H, L\}$. When the leader signs the agreement and n followers join the treaty, where $n = \{1, 2, ..., N\}$, the IEA is enacted as long as $n > 0$. When all followers sign the agreement, the leader's payoff is $V_1(S, S_N; \theta_K)$.

Like in section 3, the follower in position t obtains an equilibrium payoff of $V_t(NS; \theta_K)$ when the leader does not sign the agreement, a payoff of $V_t(S, S_n, NS_{N-n}; \theta_K)$ when the leader and $n \in N$ followers participate, and $V_t(S, S_N; \theta_K)$ when all countries sign the treaty. We maintain our previous assumption about the follower's incentives to participate in a treaty with a high-type leader but to avoid an agreement with a low-type leader. Furthermore, we consider that the global environmental quality that the t-follower obtains from joining a treaty with n signatories is positive when the leader's type is high, i.e., $V_t(S, S_{n+1}, NS_{N-(n+1)}; \theta_H) > V_t(S, S_n, NS_{N-n}; \theta_H)$, implying $BS_t((\theta_H, \theta_H), n) > 0$. In addition, follower t's incentives to participate in the agreement are weakly decreasing in the number of countries who join the treaty, i.e., $BS_t((\theta_H, \theta_H), n)$ weakly decreases in n. Intuitively, the follower's free-riding incentives increase as more countries participate in the

[12] Note that when priors satisfy $p < \overline{p}_H$ and $q < \overline{q}_L$ (or when $p \geq \overline{p}_H$ and $q < \overline{q}_L$), a separating equilibrium arises where social welfare coincides with that under complete information. We hence focus on the regions of prior probabilities for which pooling or semiseparating equilibria emerge.

IEA.[13,14] In the case that the leader's type is low, follower t obtains $V_t(S, S_{n+1}, NS_{N-(n+1)}; \theta_L)$ when it joins a group of n signatories and $V_t(S, S_n, NS_{N-n}; \theta_L)$ when it does not join, where $V_t(S, S_{n+1}, NS_{N-(n+1)}; \theta_L) < V_t(S, S_n, NS_{N-n}; \theta_L)$. That is, followers do not have incentives to participate in the agreement with a low-type leader, i.e., $BS_t((\theta_L, \theta_H), n) < 0$ for all n, which for simplicity, we consider to be constant in n.

Proposition 5. *In the IEA signaling game with N followers, only a pooling strategy profile can be sustained as a PBE in pure strategies, and it survives the Cho and Kreps' (1987) Intuitive Criterion. Specifically, the leader signs the agreement regardless of its type and any follower at position t responds signing the treaty for any* $p > \widehat{p}(k_{t-1})$, *where* k_{t-1} *denotes the number of signatory countries that follower* $t = \{1, 2, ..., N\}$ *observes participating in the agreement before itself, and* $\widehat{p}(k_{t-1}) \equiv \frac{-BS_t((\theta_L,\theta_H),k_{t-1})}{BS_t((\theta_H,\theta_H),k_{t-1})-BS_t((\theta_L,\theta_H),k_{t-1})}$.

In particular, note that the number of signatories observed by the last follower, k_{N-1}, is weakly larger than the number observed by any previous follower t, k_{t-1}, where $t \neq N$. Furthermore, since $BS_t((\theta_H, \theta_H), k_{t-1})$ decreases in k_{t-1} and $BS_t((\theta_L, \theta_H), k_{t-1})$ is constant, cutoff $\widehat{p}(k_{t-1})$ increases in the number of signatories. Intuitively, as more signatories join the treaty, followers react not signing the agreement under larger sets of parameter values, i.e., under more prior probabilities p. Thus, the increasing pattern of $\widehat{p}(k_{t-1})$ reflects the presence of free-riding incentives as more countries sign the IEA, affecting the structure of the pooling equilibrium described above. In particular, all followers participate in the IEA if and only if the prior probability is sufficiently high, i.e., $p > \widehat{p}(k_{N-1})$, all but the last follower sign the treaty if priors are moderately high, i.e., $\widehat{p}(k_{N-1}) > p > \widehat{p}(k_{N-2})$, and similarly for lower priors, where fewer followers participate in the agreement. Indeed, higher priors about the leader's type being high increase every follower's expected utility from participating in the agreement, and thus pooling equilibria with more followers signing can be sustained.[15] Finally, note that when priors are sufficiently low, i.e., $\widehat{p}(k_0) > p > \widehat{p}(k_1)$, only the first follower signs the treaty, but if priors are further reduced, i.e., $p < \widehat{p}(k_0)$, no follower participates using pure strategies and, as a result, no pooling equilibrium can be sustained. The following proposition shows that equilibria involving mixed strategies can nonetheless be supported in this region of parameter values.

Proposition 6. *A semi-separating strategy profile can be sustained as a PBE of the IEA signaling game with N followers, when* $p < \widehat{p}(k_0)$, *in which the leader signs the agreement with*

[13] This implies that the number of followers, N, is sufficiently low so that $BS_t((\theta_H, \theta_H), n) > 0$ holds for all n. Otherwise, $BS_t((\theta_H, \theta_H), n)$ could become negative, inducing all followers to not participate in the treaty irrespective of the leader's type, eliminating the potential of information transmission in the model. The proof of proposition 5 explores the special case in which the benefits from signing the agreement are constant in the number of signatories.

[14] Hence, follower t's payoff is higher when n countries participate in the IEA (but follower t does not) than when $(n-1)$ countries sign the treaty and country t also joins the IEA. That is, for a given number of followers implementing the content of the agreement, follower t's payoff is higher when it does not bear the cost of reducing emissions (signing the treaty) than when it does.

[15] Similarly, for a given prior p, an increase in the follower's incentive of joining a treaty with a high-type leader raises $BS_t((\theta_H, \theta_H), k_{t-1})$ for any follower t, lowering cutoff $\widehat{p}(k_{t-1})$, and expanding as a consequence the set of priors under which all followers participate in the agreement, all but one participate, etc.

probability $p_L \in (0,1)$ when its type is low, and signs the treaty with probability one when its type is high, $p_H = 1$; and at least one follower responds by signing the agreement with probability $r \in (0,1)$.

Our previous results hence show that for any prior probability, p, the environmental agreement is signed with positive probability by the leader and a number of followers. Specifically, the semi-separating equilibrium can be sustained when priors are sufficiently low, $p < \widehat{p}(k_0)$, whereas the pooling equilibrium where all N followers sign the IEA can be supported when priors are sufficiently high, i.e., $p > \widehat{p}(k_N)$. For intermediate priors, our results predict pooling equilibria in which a subset of followers participate in the agreement (see figure 3 below). Therefore, the introduction of several followers affects our equilibrium predictions. In particular, the set of priors sustaining the pooling equilibrium in which all countries participate in the agreement is larger with only one potential follower, i.e., $p > \overline{p}_H$, than with several followers, i.e., $p > \widehat{p}(k_N)$, since $\widehat{p}(k_N) > \overline{p}_H$. Intuitively, when several countries are involved in the negotiation, free-riding incentives become stronger, making the pooling equilibrium in which all countries sign the agreement sustainable under a more restrictive set of parameter values. Nonetheless, the set of parameter values supporting this pooling equilibrium sustains now "partially cooperative" pooling equilibria in which a subset of followers participate in the agreement. The following figure illustrates the effect of enlarging the number of followers.

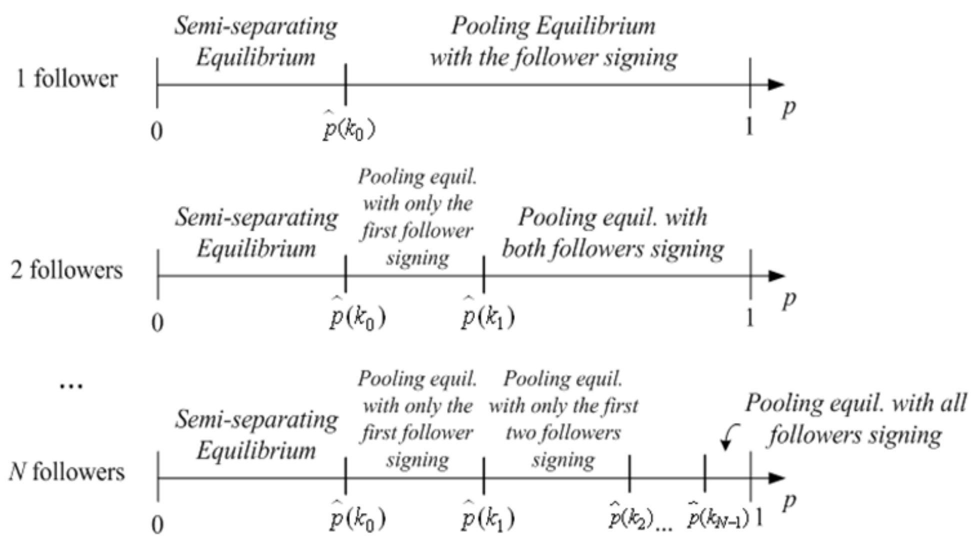

Figure 3. Equilibrium predictions for N followers.

6 Conclusions

This paper investigates the role of uncertainty as a tool to support cooperation in IEAs. We show that when negotiations take place under unilateral uncertainty treaties can become successful with positive probability, even under parameter conditions for which no agreement is reached under complete information. Specifically, we demonstrate that when priors are sufficiently high a pooling equilibrium can be sustained and, otherwise, a semiseparating equilibrium can be supported where both types of leader sign the IEA with strictly positive probabilities. These two equilibria can still be sustained after introducing an additional layer of uncertainty. Nonetheless, a separating equilibrium emerges in this context where the leader participates in the treaty only when its technology is high. We then evaluate the welfare properties of these equilibria, showing that further layers of uncertainty might enhance social welfare under certain conditions.

Our model considers that countries' technology is exogenous. In an enlarged setting, governments could invest in research and development in new technology in a previous stage of the game. For the signature decision to remain informative, those investments would have to be unobserved by other countries. This can potentially modify the ranking between equilibrium payoffs described in our paper, yielding different results.

We extend our model of unilateral uncertainty to the case of N followers, showing that equilibrium outcomes are not substantially affected. A further extension can study how the results under bilateral uncertainty are modified when N followers participate in the negotiations. Finally, another venue of further research could consider countries' lack of information about the environmental benefits from investing in clean technologies or reducing emissions since, for instance, developed countries might be more capable of assessing these benefits than underdeveloped countries are. Such asymmetric information would affect countries' equilibrium payoffs from participating in the treaty, potentially modifying our equilibrium results.

7 Appendix

7.1 *Appendix 1 - Parametric example*

Given that countries' incentives to abate share the properties of a public good game, consider a welfare function

$$\log(x_1 + x_2) + w - \frac{x_1}{\theta_K}$$

where $b(x_1, x_2) = \log(x_1 + x_2) + w$, and $c(x_1, \theta_K) = \frac{x_1}{\theta_K}$, which satisfy the properties described in section 2. When the treaty is unsuccessful, countries independently select their abatement levels in the Nash equilibrium of the game. In particular, when both the leader's and follower's technology is high, both countries have the same best response function $x_i(x_j) = \theta_H - x_j$ for all $j \neq i$. Therefore, multiple Nash equilibria exist (i.e., all pairs (x_1, x_2) satisfying $x_1 + x_2 = \theta_H$) and, for simplicity, we consider the symmetric equilibrium abatement levels $x_1(\theta_H, \theta_H) = x_2(\theta_H, \theta_H) = \frac{\theta_H}{2}$, yielding a social welfare of

$$V_1(NS; \theta_H) = V_2(NS; \theta_H) = w - \frac{1}{2} + \log(\theta_H)$$

When the leader's type is low but the follower's is high, the leader's best response function is $x_1 = \theta_L - x_2$ while that of the follower is $x_2 = \theta_H - x_1$. Hence, the best response functions only cross at the corner abatement levels $x_1(\theta_L, \theta_H) = 0$ and $x_2(\theta_L, \theta_H) = \theta_H$, entailing equilibrium welfare of

$$\begin{aligned} V_1(NS; \theta_L) &= w + \log(\theta_H) \text{ for the leader, and} \\ V_2(NS; \theta_L) &= w - 1 + \log(\theta_H) \text{ for the follower.} \end{aligned}$$

When the treaty is successful, and both leader's and follower's technology is high, the social planner solves the joint welfare maximization problem, choosing $x_1^{SO}(\theta_H, \theta_H) = x_2^{SO}(\theta_H, \theta_H) = \theta_H$, which yields equilibrium welfare of

$$\begin{aligned} V_1(S, S; \theta_H) &= w - 1 + \log(2\theta_H) - NC(\theta_H, \theta_H) \text{ for the leader, and} \\ V_2(S, S; \theta_H) &= w - 1 + \log(2\theta_H) \text{ for the follower} \end{aligned}$$

If, instead, the leader's technology is low, the social planner prescribes abatement levels $x_1^{SO}(\theta_L, \theta_H) = 0$ and $x_2^{SO}(\theta_L, \theta_H) = 2\theta_H$, entailing social welfare of

$$\begin{aligned} V_1(S, S; \theta_L) &= w + \log(2\theta_H) - NC(\theta_L, \theta_H) \text{ for the leader, and} \\ V_2(S, S; \theta_L) &= w - 2 + \log(2\theta_H) \text{ for the follower} \end{aligned}$$

Note that $V_1(S, S; \theta_H) > V_1(NS; \theta_H)$ and $V_1(S, S; \theta_L) > V_1(NS; \theta_L)$, as described in the paper, when negotiation costs satisfy $NC(\theta_K, \theta_H) < 0.19$. Finally, the follower's benefit from signing the agreement with a high-type leader is

$$BS_2(\theta_H, \theta_H) \equiv V_2(S, S; \theta_H) - V_2(S, NS; \theta_H) = 0.19$$

since $V_2(S, NS; \theta_K) = V_2(NS; \theta_K)$. If the follower faces a low-type leader, then its benefits from signing the treaty are

$$BS_2(\theta_L, \theta_H) \equiv V_2(S, S; \theta_L) - V_2(S, NS; \theta_L) = -0.31$$

which implies that the probability cutoff that supports the pooling PBE under unilateral uncertainty ($p \geq \overline{p}_H$, as described in Lemma 1) is

$$\overline{p}_H \equiv \frac{-BS_2(\theta_L, \theta_H)}{BS_2(\theta_H, \theta_H) - BS_2(\theta_L, \theta_H)} = 0.61.$$

Therefore, the semiseparating equilibrium can be sustained for all $p < 0.61$, where the low-type leader randomizes with a probability $p_L = \frac{p}{1-p} \frac{\overline{p}_H}{1-\overline{p}_H}$, where $p_L \in (0, 1)$ if $p < 0.39$.

7.2 *Proof of Lemma 1*

Separating equilibrium. Let us first show that the separating strategy profile in which the leader chooses to sign (not sign) the IEA when its technology is high (low, respectively) cannot be supported as a PBE of the signaling game. Under such strategy profile, the follower's beliefs are updated according to Bayes' rule and become $\mu(H|S) = 1$ and $\mu(H|NS) = 0$. Given these posterior beliefs, the follower signs the agreement since $V_2(S, S; \theta_H) > V_2(S, NS; \theta_H)$. As a consequence, the leader chooses to participate in the treaty both when its technology is high and low, since $V_1(S, S; \theta_K) > V_1(NS; \theta_K)$ for both θ_H and θ_L. But this strategy profile for the leader contradicts the separating strategy described above, and therefore it cannot be sustained as a PBE of the game.

Pooling equilibrium. Let us next demonstrate that the pooling strategy profile in which the leader signs the IEA regardless of its type can be part of a PBE under certain conditions. In this strategy, the follower's posterior beliefs cannot be updated and thus coincide with its priors, i.e., $\mu(H|S) = p$ and $\mu(L|S) = 1 - p$. (Note that off-the-equilibrium beliefs do not play a role in this pooling equilibrium. In particular, after observing the off-the-equilibrium action of "no signature" the follower has, by definition, an empty action space. Therefore, off-the-equilibrium beliefs cannot affect the follower's response and, as a consequence, do not affect the leader's decision either.)

Given these beliefs, the follower chooses to not sign the agreement if

$$p \times V_2(S, S; \theta_H) + (1 - p) \times V_2(S, S; \theta_L) < p \times V_2(S, NS; \theta_H) + (1 - p) \times V_2(S, NS; \theta_L), \text{ or}$$

$$p < \frac{V_2(S, NS; \theta_L) - V_2(S, S; \theta_L)}{[V_2(S, S; \theta_H) - V_2(S, NS; \theta_H)] - [V_2(S, S; \theta_L) - V_2(S, NS; \theta_L)]}$$

or, more compactly, $p < \frac{-BS_2(\theta_L,\theta_H)>0}{BS_2(\theta_H,\theta_H)-BS_2(\theta_L,\theta_H)} \equiv \overline{p}_H$ where $\overline{p}_H > 0$ since the follower's payoffs satisfy $BS_2(\theta_L, \theta_H) < 0$ and $BS_2(\theta_H, \theta_H) > BS_2(\theta_L, \theta_H)$ by definition, and $\overline{p}_H < 1$ given that $BS_2(\theta_H, \theta_H) - BS_2(\theta_L, \theta_H) > -BS_2(\theta_L, \theta_H)$ or $BS_2(\theta_H, \theta_H) > 0$. Let us next analyze equilibrium strategies for the leader, for different priors p.

1. *Low priors,* $p < \overline{p}_H$. In this case the follower responds not signing the treaty. Since $V_1(S, NS; \theta_K) < V_1(NS; \theta_K)$ for all K-type leader, the leader does not participates in the treaty; which contradicts the pooling strategy profile. Thus, this pooling strategy profile cannot be supported as a PBE of the game when $p < \overline{p}_H$.

2. *High priors,* $p \geq \overline{p}_H$. In this case the follower responds signing the treaty. Since $V_1(S, S; \theta_K) > V_1(NS; \theta_K)$ for all K-type leader, the leader participates in the agreement both when its type is high and low. Hence, the pooling strategy profile can be supported as PBE of the game when $p \geq \overline{p}_H$.

Finally, let us show that the pooling strategy profile where the leader does not sign the IEA regardless of its type cannot be sustained as part of a PBE. First, note that the follower's off-the-equilibrium beliefs cannot be updated using Bayes' rule, and hence must be arbitrarily specified,

$\mu(H|S) = 1$. Given these beliefs, the follower signs the agreement. Therefore, the leader prefers to participate in the IEA, both when its type is high and low since $V_1(S,S;\theta_K) > V_1(NS;\theta_K)$ for both θ_H and θ_L. This leader's strategy, however, contradicts the prescribed equilibrium in which the leader does not sign the treaty regardless of its type, and hence this pooling strategy profile cannot be supported in this case.

Intuitive Criterion. Let us apply the Cho and Kreps' (1987) Intuitive Criterion for the pooling PBE where $p \geq \overline{p}_H$. We first check if a deviation towards "not sign" is equilibrium dominated for either type of leader. When the leader's technology is high, the highest payoff that it obtains by deviating towards "not sign" is $V_1(NS;\theta_H)$ which does not exceed its equilibrium payoff $V_1(S,S;\theta_H)$. Similarly, the highest payoff that the low-technology leader can obtain is $V_1(NS;\theta_L)$ which does not exceed its equilibrium payoff of $V_1(S,S;\theta_L)$. Hence, no type of leader has incentives to deviate towards "not sign." As a consequence, the follower's posterior beliefs cannot be updated after observing the off-the-equilibrium message "not sign", and hence the pooling PBE survives the Intuitive Criterion. ■

7.3 *Proof of Proposition 1*

Let us first analyze the follower's strategy. This player must be mixing. Otherwise, the leader could anticipate its action and play pure strategies as in any of the strategy profiles described in lemma 1, which are not PBE of the signaling game when $p < \overline{p}_H$. Hence, the follower must be indifferent between signing and not signing the treaty, that is,

$$\begin{aligned} & \mu(H|S) \times V_2(S,S;\theta_H) + (1-\mu(H|S)) \times V_2(S,S;\theta_L) \\ = \quad & \mu(H|S) \times V_2(S,NS;\theta_H) + (1-\mu(H|S)) \times V_2(S,NS;\theta_L), \end{aligned}$$

or $\mu(H|S) = \overline{p}_H$. We can next use the follower's posterior beliefs in order to find the probability with which the leader randomizes when its technology is low, p_L, by using Bayes' rule.

$$\mu(H|S) = \overline{p}_H = \frac{(1-p) \times p_L}{((1-p) \times p_L) + (p \times p_H)}$$

where $p_H = 1$. Solving for p_L, we obtain $p_L = \frac{p}{1-p}\frac{\overline{p}_H}{1-\overline{p}_H}$, which is positive, $p_L > 0$, since $p, \overline{p}_H \in (0,1)$. In addition, $p_L < 1$ for all probability p satisfying $p < 1 - \overline{p}_H$. Since $|BS_2(\theta_H,\theta_H)| > |BS_2(\theta_L,\theta_H)|$, cutoff $\overline{p}_H$ satisfies $\overline{p}_H < 1/2$, which implies $p < 1 - \overline{p}_H$ is satisfied for all priors, because $p < \overline{p}_H$ by assumption.

Finally, note that if the leader mixes with probability $p_L \in (0,1)$, it must be that the follower makes it indifferent between signing and not signing the agreement.

$$r \times V_1(S,S;\theta_L) + (1-r) \times V_1(S,NS;\theta_L) = V_1(NS;\theta_L), \text{ or}$$

$$r = \frac{V_1(NS;\theta_L) - V_1(S,NS;\theta_L)}{V_1(S,S;\theta_L) - V_1(S,NS;\theta_L)} \equiv \widehat{r}$$

where $\widehat{r}$ denotes the probability with which the follower mixes between signing and not signing the treaty. This probability is positive, $\widehat{r} > 0$, given that the low-technology leader's payoffs satisfy $V_1(NS;\theta_L) > V_1(S,NS;\theta_L)$ and $V_1(S,S;\theta_L) > V_1(S,NS;\theta_L)$, and $\widehat{r} < 1$ since $V_1(NS;\theta_L) < V_1(S,S;\theta_L)$. ■

7.4 *Proof of Proposition 2*

Separating equilibrium. Let us first show that the separating strategy profile in which the leader chooses to sign (not sign) the IEA when its technology is high (low, respectively) can be supported as a PBE of the signaling game. Under such strategy profile, the J-type follower's beliefs are updated according to Bayes' rule and become $\mu_J(H|S) = 1$ and $\mu_J(H|NS) = 0$ after observing a signature (not signature, respectively) from the leader, where $J = \{H, L\}$. Given these posterior beliefs, the follower signs the agreement since $V_2(S,S;(\theta_H,\theta_H)) > V_2(S,NS;(\theta_H,\theta_H))$ when the follower is a high-type, and similarly for the low-type follower where $V_2(S,S;(\theta_H,\theta_L)) > V_2(S,NS;(\theta_H,\theta_L))$. As a consequence, the high-type leader chooses to participate in the treaty since $V_1(S,S;(\theta_H,\theta_J)) > V_1(NS;(\theta_H,\theta_J))$ for all J-type follower. On the other hand, the low-type leader does not participate (as prescribed in this separating equilibrium) if

$$q\times V_1(S,S;(\theta_L,\theta_H))+(1-q)\times V_1(S,S;(\theta_L,\theta_L)) < q\times V_1(NS;(\theta_L,\theta_H))+(1-q)\times V_1(NS;(\theta_L,\theta_L)), \text{ or}$$

$$q < \frac{-BS_1(\theta_L,\theta_L)}{BS_1(\theta_L,\theta_H) - BS_1(\theta_L,\theta_L)} \equiv \overline{q}_L$$

where $BS_1(\theta_L,\theta_J) \equiv V_1(S,S;(\theta_L,\theta_J)) - V_1(NS;(\theta_L,\theta_J))$ denotes the low-type leader's benefit of signing an agreement with a J-type follower. Note that cutoff $\overline{q}_L > 0$ since $BS_1(\theta_L,\theta_L) < 0$ and $\overline{q}_L < 1$ given that $BS_1(\theta_L,\theta_H) > 0$ and $BS_1(\theta_L,\theta_H) > BS_1(\theta_L,\theta_L)$. Hence, the low-type leader does not participate if $q < \overline{q}_L$. Therefore, the above separating strategy profile can be sustained as a PBE of the game if $q < \overline{q}_L$.

Pooling equilibrium with signature. Let us next demonstrate that the pooling strategy profile in which the leader signs the IEA regardless of its type can be part of a PBE under certain conditions. In this strategy, the J-type follower's posterior beliefs cannot be updated and thus coincide with its priors, i.e., $\mu_J(H|S) = p$ and $\mu_J(L|S) = 1-p$. (Note that off-the-equilibrium beliefs do not play a role in this pooling equilibrium. In particular, after observing the off-the-equilibrium action of "no signature" the follower has, by definition, an empty action space. Therefore, off-the-equilibrium beliefs cannot affect the follower's response and, as a consequence, do not affect the leader's decision either.)

Given these beliefs, the high-type follower chooses to not sign the agreement if

$$\begin{aligned} & p\times V_2(S,S;(\theta_H,\theta_H)) + (1-p)\times V_2(S,S;(\theta_L,\theta_H)) \\ < \; & p\times V_2(S,NS;(\theta_H,\theta_H)) + (1-p)\times V_2(S,NS;(\theta_L,\theta_H)), \text{ or} \end{aligned}$$

$$p < \frac{-BS_2\left(\theta_L, \theta_H\right)}{BS_2\left(\theta_H, \theta_H\right) - BS_2\left(\theta_L, \theta_H\right)} \equiv \overline{p}_H$$

where $BS_2\left(\theta_K, \theta_H\right) \equiv V_2(S, S; (\theta_K, \theta_H)) - V_2(S, NS; (\theta_K, \theta_H))$ denotes the high-type follower's benefit from signing an agreement with a K-type leader. Note that $\overline{p}_H > 0$ since the follower's payoffs satisfy $BS_2\left(\theta_L, \theta_H\right) < 0$ and $BS_2\left(\theta_H, \theta_H\right) > BS_2\left(\theta_L, \theta_H\right)$ by definition. In addition, $\overline{p}_H < 1$ given that $BS_2\left(\theta_H, \theta_H\right) - BS_2\left(\theta_L, \theta_H\right) > -BS_2\left(\theta_L, \theta_H\right)$ or $BS_2\left(\theta_H, \theta_H\right) > 0$. Hence, when $p < \overline{p}_H$, the high-type follower does not sign the treaty, and signs otherwise. Similarly, the low-type follower does not sign the agreement if

$$\begin{aligned} & p \times V_2(S, S; (\theta_H, \theta_L)) + (1-p) \times V_2(S, S; (\theta_L, \theta_L)) \\ < \quad & p \times V_2(S, NS; (\theta_H, \theta_L)) + (1-p) \times V_2(S, NS; (\theta_L, \theta_L)), \text{ or} \end{aligned}$$

$$p < \frac{-BS_2\left(\theta_L, \theta_L\right)}{BS_2\left(\theta_H, \theta_L\right) - BS_2\left(\theta_L, \theta_L\right)} \equiv \overline{p}_L$$

where $BS_2\left(\theta_K, \theta_L\right) \equiv V_2(S, S; (\theta_K, \theta_L)) - V_2(S, NS; (\theta_K, \theta_L))$ denotes the low-type follower's benefit from signing an agreement with a K-type leader. Since $BS_2\left(\theta_L, \theta_L\right) \geq 0$ by definition, $\overline{p}_L \leq 0$, implying that all priors $p \in [0, 1]$ satisfy $p \geq \overline{p}_L$, and therefore the low-type follower signs for all parameter values. Let us next analyze equilibrium strategies for the leader.

1. *High priors,* $p \geq \overline{p}_H$. In this case both types of follower respond signing the treaty. The high-type leader participates in the agreement since $V_1(S, S; (\theta_H, \theta_J)) > V_1(NS; (\theta_H, \theta_J))$ for all follower J. However, the low-type leader signs if

$$\begin{aligned} & q \times V_1(S, S; (\theta_L, \theta_H)) + (1-q) \times V_1(S, S; (\theta_L, \theta_L)) \\ \geq \quad & q \times V_1(NS; (\theta_L, \theta_H)) + (1-q) \times V_1(NS; (\theta_L, \theta_L)), \text{ or} \end{aligned}$$

$$q \geq \frac{-BS_1(\theta_L, \theta_L)}{BS_1(\theta_L, \theta_H) - BS_1(\theta_L, \theta_L)} \equiv \overline{q}_L$$

where $\overline{q}_L \in (0, 1)$ from our above discussion in the separating equilibrium. Therefore, the pooling strategy profile in which both types of leader sign the agreement can be sustained if $p \geq \overline{p}_H$ and $q \geq \overline{q}_L$.

2. *Low priors,* $p < \overline{p}_H$. In this case the high-type follower responds not participating in the treaty while the low-type follower signs for all parameter values. The high-type leader participates in the agreement for all priors q since $V_1(S, NS; (\theta_H, \theta_H)) > V_1(NS; (\theta_H, \theta_H))$ when facing a high-type follower and $V_1(S, S; (\theta_H, \theta_L)) > V_1(NS; (\theta_H, \theta_L))$ when facing a low-type follower. Regarding the low-type leader, he signs the treaty if

$$\begin{aligned} & qV_1(S, NS; (\theta_L, \theta_H)) + (1-q)V_1(S, S; (\theta_L, \theta_L)) \\ \geq \quad & qV_1(NS; (\theta_L, \theta_H)) + (1-q)V_1(NS; (\theta_L, \theta_L)), \text{ or} \end{aligned}$$

$$q \geq \frac{-BS_1(\theta_L,\theta_L)}{[V_1(S,NS;(\theta_L,\theta_H)) - V_1(NS;(\theta_L,\theta_H))] - BS_1(\theta_L,\theta_L)} \equiv \overline{q}'_L$$

where $BS_1(\theta_L,\theta_L) < 0$ by definition. In addition, $V_1(S,NS;(\theta_L,\theta_H)) < V_1(NS;(\theta_L,\theta_H))$ since the agreement is not successful. Hence, cutoff $\overline{q}'_L > 1$ and the low-type leader does not sign the agreement for any prior q. Therefore, the pooling strategy profile in which both types of leader sign the treaty cannot be sustained as PBE when $p < \overline{p}_H$.

Pooling equilibrium with no signature. Finally, let us show that the pooling strategy profile where the leader does not sign the IEA regardless of its type cannot be sustained as part of a PBE. First, note that the follower's posterior beliefs cannot be updated using Bayes' rule, and hence must be arbitrarily specified, $\mu_J(H|S) = 1$ for any J-type follower. Given these beliefs, both the high- and low-type followers respond signing the agreement. Hence, when the leader is a high type, it signs the agreement since $V_1(S,S;(\theta_H,\theta_J)) > V_1(NS;(\theta_H,\theta_J))$ for any J-type follower. Therefore, the high-type leader signs the treaty under all priors q and the pooling strategy profile in which no leader signs cannot be supported as PBE.

Intuitive Criterion. Let us apply the Cho and Kreps' (1987) Intuitive Criterion for the pooling PBE where $p \geq \overline{p}_H$ and $q \geq \overline{q}_L$. We first check if a deviation towards "not sign" is equilibrium dominated for either type of leader. When the leader is a high type, the highest payoff that it obtains by deviating towards "not sign" is $qV_1(NS;(\theta_H,\theta_H)) + (1-q)V_1(NS;(\theta_H,\theta_L))$, which does exceed its equilibrium payoff, $qV_1(S,S;(\theta_H,\theta_H)) + (1-q)V_1(S,S;(\theta_H,\theta_L))$, since $V_1(S,S;(\theta_H,\theta_J)) > V_1(NS;(\theta_H,\theta_J))$ for all J. Regarding the low-type leader, the highest payoff that it can obtain by deviating is $qV_1(NS;(\theta_L,\theta_H)) + (1-q)V_1(NS;(\theta_L,\theta_L))$ which exceeds its equilibrium payoff of $qV_1(S,S;(\theta_L,\theta_H)) + (1-q)V_1(S,S;(\theta_L,\theta_L))$ if

$$\begin{aligned} & qV_1(S,S;(\theta_L,\theta_H)) + (1-q)V_1(S,S;(\theta_L,\theta_L)) \\ < \; & qV_1(NS;(\theta_L,\theta_H)) + (1-q)V_1(NS;(\theta_L,\theta_L)), \text{ or } q < \overline{q}_L \end{aligned}$$

where cutoff $\overline{q}_L \in (0,1)$ from our above discussion in the separating equilibrium. Hence, the low-type leader deviates towards "not sign" if $q < \overline{q}_L$. Therefore, the pooling equilibrium in which both types of leader do not participate in the treaty (supported under $p \geq \overline{p}_H$ and $q \geq \overline{q}_L$) survives the Cho and Kreps' (1987) Intuitive Criterion. ■

7.5 *Proof of Proposition 3*

Let us first analyze the strategy for the high-type follower. (The low-type follower signs the agreement for all priors p, and therefore it does not modify its signature decision based on the information inferred from the leader's randomization).The high-type follower must be mixing. Otherwise, the leader could anticipate its action and play pure strategies as in any of the strategy profiles described in proposition 2, which are not PBE of the signaling game when $p < \overline{p}_H$ and $q \geq \overline{q}_L$. Hence, the high-type follower must be indifferent between signing and not signing the treaty, that is,

$$\begin{aligned} & \mu_H(H|S) \times V_2(S, S; (\theta_H, \theta_H)) + (1 - \mu_H(H|S)) \times V_2(S, S; (\theta_L, \theta_H)) \\ = \; & \mu_H(H|S) \times V_2(S, NS; (\theta_H, \theta_H)) + (1 - \mu_H(H|S)) \times V_2(S, NS; (\theta_L, \theta_H)), \end{aligned}$$

or $\mu_H(H|S) = \overline{p}_H$. We can next use the follower's posterior beliefs in order to find the probability with which the leader randomizes when its type is low, p_L, by using Bayes' rule.

$$\mu_H(H|S) = \overline{p}_H = \frac{(1-p) \times p_L}{((1-p) \times p_L) + (p \times p_H)}$$

where $p_H = 1$. Solving for p_L, we obtain $p_L = \frac{p}{1-p} \frac{\overline{p}_H}{1-\overline{p}_H}$, which is positive, $p_L > 0$, since $p, \overline{p}_H \in (0, 1)$. In addition, note that $p_L < 1$ holds for all priors p satisfying $p < 1 - \overline{p}_H$. Since $|BS_2(\theta_H, \theta_H)| > |BS_2(\theta_L, \theta_H)|$ by definition, cutoff $\overline{p}_H$ satisfies $\overline{p}_H < 1/2$, which implies $p < 1 - \overline{p}_H$ holds for all priors, because $p < \overline{p}_H$ by assumption. Note that probability p_L increases in p.

Finally, note that if the low-type leader mixes with probability $p_L \in (0, 1)$, it must be that the high-type follower makes this leader indifferent between signing and not signing the agreement (the low-type follower responds by signing under all parameter conditions). Using $r_H(q)$ to denote the probability with which the high-type follower mixes between signing and not signing the treaty, the low-type leader is indifferent if

$$\begin{aligned} & q\left[r_H(q) \times V_1(S, S; (\theta_L, \theta_H)) + (1 - r_H(q)) \times V_1(S, NS; (\theta_L, \theta_H))\right] + (1-q) V_1(S, S; (\theta_L, \theta_L)) \\ = \; & q V_1(NS; (\theta_L, \theta_H)) + (1-q) V_1(NS; (\theta_L, \theta_L)), \text{ or} \end{aligned}$$

$$r_H(q) = \frac{NC(\theta_L, \theta_H)}{BPS_1(\theta_L, \theta_H)} + \frac{1-q}{q} \frac{-BS_1(\theta_L, \theta_L)}{BPS_1(\theta_L, \theta_H)}$$

where $BPS_1(\theta_L, \theta_J) \equiv V_1(S, S; (\theta_L, \theta_J)) - V_1(S, NS; (\theta_L, \theta_J))$ denotes the benefits that the low-type leader obtains from the posterior signature of the agreement by the J-type follower. On the other hand, $NC(\theta_L, \theta_J) \equiv V_1(NS; (\theta_L, \theta_J)) - V_1(S, NS; (\theta_L, \theta_J))$ represents the low-type leader's negotiation costs. It is easy to show that this leader's benefit from signing the treaty, $BS_1(\theta_L, \theta_J)$, can therefore be expressed as the sum of the above two benefits, i.e., $BS_1(\theta_L, \theta_J) = BPS_1(\theta_L, \theta_J) - NC(\theta_L, \theta_J)$. Note that for the specific case in which $q = 1$, $r_H(q)$ becomes $r_H(1) = \frac{NC(\theta_L, \theta_H)}{BPS_1(\theta_L, \theta_H)}$, where this probability satisfies $r_H(1) > 0$ since $NC(\theta_L, \theta_H) > 0$ and $BPS_1(\theta_L, \theta_H) > 0$. In addition, $r_H(1) < 1$ given that $NC(\theta_L, \theta_J) < BPS_1(\theta_L, \theta_H)$ provided that condition

$$V_1(NS; (\theta_L, \theta_H)) - V_1(S, NS; (\theta_L, \theta_H)) < V_1(S, S; (\theta_L, \theta_H)) - V_1(S, NS; (\theta_L, \theta_H))$$

holds since $V_1(NS; (\theta_L, \theta_H)) < V_1(S, S; (\theta_L, \theta_H))$ by definition. Generally, for any prior q, the probability $r_H(q)$ satisfies $r_H(q) \in (0, 1)$. Indeed, note that $NC(\theta_L, \theta_H) + \frac{1-q}{q}[-BS_1(\theta_L, \theta_L)] < BPS_1(\theta_L, \theta_H)$ implies $\frac{1-q}{q} < \frac{BS_1(\theta_L, \theta_H)}{-BS_1(\theta_L, \theta_L)}$, which can be rearranged as $\overline{q}_L \equiv \frac{-BS_1(\theta_L, \theta_L)}{BS_1(\theta_L, \theta_H) - BS_1(\theta_L, \theta_L)} < q$, which holds by assumption. Finally, note that probability $r_H(q)$ is decreasing in q. ■

7.6 *Proof of Proposition 4*

Separating PBE. When the leader's technology is high, it participates in the agreement and the follower responds joining, yielding a social welfare (summing up the equilibrium payoffs of leader and follower) of $V_1(S, S; (\theta_H, \theta_J)) + V_2(S, S; (\theta_H, \theta_J))$ where $J = \{H, L\}$ denotes the follower's type. If, in contrast, the leader's technology is low, it does not sign the treaty, entailing a welfare of $V_1(NS; (\theta_L, \theta_J)) + V_2(NS; (\theta_L, \theta_J))$.

Pooling PBE. The pooling PBE yields a social welfare of $V_1(S, S; (\theta_K, \theta_J)) + V_2(S, S; (\theta_K, \theta_J))$, where $K = \{H, L\}$ denotes the leader's type, which entails the same social welfare as in the separating PBE when the leader's type is high, i.e., $SW_{separ} = SW_{pooling}$, for any follower's type J. However, when the leader's type is low, the separating equilibrium prescribes that this leader does not participate in the treaty, yielding a social welfare of $V_1(NS; (\theta_L, \theta_J)) + V_2(NS; (\theta_L, \theta_J))$, which lies weakly below that under the pooling PBE if

$$V_1(NS; (\theta_L, \theta_J)) + V_2(NS; (\theta_L, \theta_J)) \leq V_1(S, S; (\theta_L, \theta_J)) + V_2(S, S; (\theta_L, \theta_J))$$

or alternatively, $-BS_2(\theta_L, \theta_J) \leq BS_1(\theta_L, \theta_J)$, where $BS_1(\theta_L, \theta_J) \equiv V_1(S, S; (\theta_L, \theta_J)) - V_1(NS; (\theta_L, \theta_J))$ denotes the low-type leader's benefit from signing an agreement with a J-type follower, and conversely $BS_2(\theta_L, \theta_J) \equiv V_2(S, S; (\theta_L, \theta_J)) - V_2(S, NS; (\theta_L, \theta_J))$ represents the J-type follower's benefit from signing a treaty with a low-type leader, since $V_2(NS; (\theta_L, \theta_J)) = V_2(S, NS; (\theta_L, \theta_J))$. Let us separately analyze the cases in which the follower's type is high and low.

1. If the leader's technology is low while that of the follower is high, i.e., $K = L$ and $J = H$, then $BS_1(\theta_L, \theta_H) > 0$ for the leader and $BS_2(\theta_L, \theta_H) < 0$ for the follower and, as a consequence, condition $-BS_2(\theta_L, \theta_H) \leq BS_1(\theta_L, \theta_H)$ holds if

$$V_1(S, S; (\theta_L, \theta_H)) + V_2(S, S; (\theta_L, \theta_H)) > V_1(NS; (\theta_L, \theta_H)) + V_2(NS; (\theta_L, \theta_H)),$$

which implies

$$\begin{aligned} & 2 \times \left[b(x_1^{SO}(\theta_L, \theta_H), x_2^{SO}(\theta_L, \theta_H)) - b(x_1(\theta_L, \theta_H), x_2(\theta_L, \theta_H))\right] \\ > \; & \left[c(x_1^{SO}(\theta_L, \theta_H), \theta_L) - c(x_1(\theta_L, \theta_H), \theta_L)\right] + \left[c(x_2^{SO}(\theta_L, \theta_H), \theta_H) - c(x_2(\theta_L, \theta_H), \theta_H)\right] + NC(\theta_L, \theta_H) \end{aligned}$$

which represents that the environmental benefits from signing the treaty are sufficiently high, implying that $SW_{separ} < SW_{pooling}$. Otherwise, $-BS_2(\theta_L, \theta_H) > BS_1(\theta_L, \theta_H)$, and the separating equilibrium yields a larger social welfare than the pooling equilibrium.

2. When the leader's and follower's technology are low, i.e., $K, J = L$, the benefits from signing an agreement between two low-type countries are $BS_1(\theta_L, \theta_L) < 0$ for the leader and $BS_2(\theta_L, \theta_L) \geq 0$ for the follower. Therefore, given the symmetry in benefit and cost functions

among two low-type countries, $BS_1(\theta_L,\theta_L) = BS_2(\theta_L,\theta_L) - NC(\theta_L,\theta_L)$, implying that condition $-BS_2(\theta_L,\theta_L) \leq BS_1(\theta_L,\theta_L)$ does not hold if $-BS_2(\theta_L,\theta_L) > BS_2(\theta_L,\theta_L) - NC(\theta_L,\theta_L)$ or $NC(\theta_L,\theta_L) > 2BS_2(\theta_L,\theta_L)$, hence social welfare in the separating equilibrium is larger than under the pooling equilibrium, i.e., $SW_{separ} > SW_{pooling}$. Otherwise, $SW_{separ} < SW_{pooling}$.

Semiseparating PBE. Let us now evaluate social welfare in the semiseparating equilibrium.

1. When both the leader and follower's types are high, the former participates with probability one while the latter randomizes according to probability $r_H(q)$. Hence, social welfare becomes

$$r_H(q) \times [V_1(S,S;(\theta_H,\theta_H)) + V_2(S,S;(\theta_H,\theta_H))]$$
$$+(1-r_H(q)) \times [V_1(S,NS;(\theta_H,\theta_H)) + V_2(S,NS;(\theta_H,\theta_H))]$$

where the first term in brackets, $V_1(S,S;(\theta_H,\theta_H)) + V_2(S,S;(\theta_H,\theta_H))$, is larger than the second, $V_1(S,NS;(\theta_H,\theta_H)) + V_2(S,NS;(\theta_H,\theta_H))$, since $V_1(S,S;(\theta_H,\theta_H)) > V_1(S,NS;(\theta_H,\theta_H))$ for the leader and similarly $V_2(S,S;(\theta_H,\theta_H)) > V_2(S,NS;(\theta_H,\theta_H))$ for the follower, since $BS_2(\theta_H,\theta_H) > 0$. Furthermore, recall that social welfare under the pooling PBE is

$$V_1(S,S;(\theta_H,\theta_H)) + V_2(S,S;(\theta_H,\theta_H)).$$

Hence, welfare in the semiseparating equilibrium is a linear combination between the social welfare in the pooling PBE and a smaller number, thereby yielding a lower welfare than under the pooling PBE. Combining this result with that from the pooling PBE, we obtain that $SW_{separ} = SW_{pooling} > SW_{semisepar}$.

1. When the leader's type is high but the follower's is low, this equilibrium prescribes that both countries sign the agreement with probability one, thus yielding a social welfare of $V_1(S,S;(\theta_H,\theta_L)) + V_2(S,S;(\theta_H,\theta_L))$, which coincides with that under the pooling equilibrium of Proposition 1b. Combining this result with that from the pooling PBE, we obtain that $SW_{separ} = SW_{pooling} = SW_{semisepar}$.

2. Finally, when leader's type is low and that of the follower is high, both countries randomize their participation decision, yielding social welfare of

$$p_L r_H(q)\,[V_1(S,S;(\theta_L,\theta_H)) + V_2(S,S;(\theta_L,\theta_H))]$$
$$+p_L(1-r_H(q))\,[V_1(S,NS;(\theta_L,\theta_H)) + V_2(S,NS;(\theta_L,\theta_H))]$$
$$+(1-p_L)\,[V_1(NS;(\theta_L,\theta_H)) + V_2(NS;(\theta_L,\theta_H))]\,.$$

whereas social welfare under the pooling PBE is $V_1(S,S;(\theta_L,\theta_H)) + V_2(S,S;(\theta_L,\theta_H))$, which we denote as A. Hence, the social welfare in the semiseparating equilibrium is lower than under pooling PBE if

$$p_L \left[r_H(q)A + (1-r_H(q))\left[V_1(S,NS;(\theta_L,\theta_H)) + V_2(S,NS;(\theta_L,\theta_H))\right]\right]$$
$$+(1-p_L)\left[V_1(NS;(\theta_L,\theta_H)) + V_2(NS;(\theta_L,\theta_H))\right] < A$$

rearranging, using the property that $V_2(S,NS;(\theta_L,\theta_H)) = V_2(NS;(\theta_L,\theta_H)) = X$, and solving for the payoff A, we obtain

$$\overline{A} \equiv \frac{-p_L NC(\theta_L,\theta_H) + B - p_L r_H(q)\left[V_1(S,NS;(\theta_L,\theta_H)) + X\right]}{1 - p_L r_H(q)} < A$$

where $NC(\theta_L,\theta_H) \equiv V_1(NS;(\theta_L,\theta_H)) - V_1(S,NS;(\theta_L,\theta_H))$, and $B \equiv V_1(NS;(\theta_L,\theta_H)) + V_2(NS;(\theta_L,\theta_H))$. Cutoff $\overline{A}$ lies below B, which implies that $A > B$ is a sufficient condition for $A > \overline{A}$, entailing that welfare in the pooling equilibrium exceeds that in the semiseparating equilibrium. In particular, $\overline{A} < B$ since

$$-p_L NC(\theta_L,\theta_H) + B - p_L r_H(q)\left[V_1(S,NS;(\theta_L,\theta_H)) + X\right] < B\left[1 - p_L r_H(q)\right]$$

which implies

$$\frac{NC(\theta_L,\theta_H)}{r_H(q)} + \left[V_1(S,NS;(\theta_L,\theta_H)) + X\right] > B$$

Using $B \equiv V_1(NS;(\theta_L,\theta_H)) + V_2(NS;(\theta_L,\theta_H)) = V_1(NS;(\theta_L,\theta_H)) + X$, since $V_2(NS;(\theta_L,\theta_H)) = X$, the above inequality can be simplified to $NC(\theta_L,\theta_H) > r_H(q)NC(\theta_L,\theta_H)$, which holds by assumption given that $r_H(q) \in (0,1)$. Concluding, condition $A > B$, or alternatively,

$$V_1(S,S;(\theta_L,\theta_H)) + V_2(S,S;(\theta_L,\theta_H)) > V_1(NS;(\theta_L,\theta_H)) + V_2(NS;(\theta_L,\theta_H))$$

guarantees that welfare in the pooling equilibrium exceeds that in the semiseparating equilibrium.

- Finally, in order to obtain a complete welfare ranking, let us now compare welfare under the separating and semiseparating equilibrium when the leader's type is low and the follower's is high. In the separating PBE, social welfare is $V_1(NS;(\theta_L,\theta_H)) + V_2(NS;(\theta_L,\theta_H)) \equiv B$. The semiseparating PBE yields a welfare of

$$p_L \left[r_H(q)A + (1-r_H(q))\left[V_1(S,NS;(\theta_L,\theta_H)) + X\right]\right] + (1-p_L)\left[V_1(NS;(\theta_L,\theta_H)) + X\right],$$

which is larger than in the separating equilibrium, B, if

$$A > \frac{p_L NC(\theta_L,\theta_H) - (1-p_L)V_1(S,NS;(\theta_L,\theta_H)) + r_H(q)\left[V_1(S,NS;(\theta_L,\theta_H)) + X\right]}{p_L r_H(q)} \equiv \widetilde{A}$$

Hence, if $A > \widetilde{A}$ (i.e., the benefits from signing the agreement are sufficiently high) we obtain the complete welfare ranking $SW_{pooling} > SW_{semisepar} > SW_{separ}$. If, instead, $A \leq \widetilde{A}$, the welfare ranking becomes $SW_{pooling} > SW_{separ} > SW_{semisepar}$.

3. When, in contrast, both the leader and follower's types are low, the former randomizes according to a probability $p_L = \frac{p}{1-p}\frac{\overline{p}_H}{1-\overline{p}_H}$, whereas the follower responds joining the treaty with probability one. Therefore, social welfare in this case is

$$p_L \times [V_1(S,S;(\theta_L,\theta_L)) + V_2(S,S;(\theta_L,\theta_L))] + (1-p_L) \times [V_1(NS;(\theta_L,\theta_L)) + V_2(NS;(\theta_L,\theta_L))].$$

which can be alternatively expressed as

$$\begin{aligned}&p_L \times [V_1(S,S;(\theta_L,\theta_L)) - V_1(NS;(\theta_L,\theta_L))]\\&+p_L \times [V_2(S,S;(\theta_L,\theta_L)) - V_2(NS;(\theta_L,\theta_L))]\\&+V_1(NS;(\theta_L,\theta_L)) + V_2(NS;(\theta_L,\theta_L)).\end{aligned}$$

where the first term in brackets is negative since $V_1(S,S;(\theta_L,\theta_L)) < V_1(NS;(\theta_L,\theta_L))$ for the leader, given that $BS_1(\theta_L,\theta_L) < 0$ by assumption. The second term is weakly positive because $V_2(S,S;(\theta_L,\theta_L)) \geq V_2(NS;(\theta_L,\theta_L))$ for the follower, which lies above the welfare under the pooling PBE, $V_1(S,S;(\theta_L,\theta_L)) + V_2(S,S;(\theta_L,\theta_L))$, if

$$p_L \times [V_1(S,S;(\theta_L,\theta_L)) - V_1(NS;(\theta_L,\theta_L))] + p_L \times [V_2(S,S;(\theta_L,\theta_L)) - V_2(NS;(\theta_L,\theta_L))]$$

$$+V_1(NS;(\theta_L,\theta_L)) + V_2(NS;(\theta_L,\theta_L)) > V_1(S,S;(\theta_L,\theta_L)) + V_2(S,S;(\theta_L,\theta_L))$$

which can be expressed as

$$-BS_1(\theta_L,\theta_L)(1+p_L) > BS_2(\theta_L,\theta_L)(1-p_L)$$

if negotiation costs satisfy $NC(\theta_L,\theta_H) > 2BS_2(\theta_L,\theta_L)$ then $-BS_2(\theta_L,\theta_L) > BS_1(\theta_L,\theta_L)$ holds. Therefore, $BS_2(\theta_L,\theta_L) < -BS_1(\theta_L,\theta_L)$,which implies that $-BS_1(\theta_L,\theta_L)(1+p_L) > BS_2(\theta_L,\theta_L)(1-p_L)$ is satisfied, and hence $SW_{semisepar} > SW_{pooling}$.

- Finally, in order to obtain a complete welfare ranking, let us now compare social welfare under the separating and semiseparating equilibrium. In the separating PBE, social welfare is $V_1(NS;(\theta_L,\theta_L))+V_2(NS;(\theta_L,\theta_L))$; whereas in the semiseparating PBE social welfare is

$$p_L \times BS_1(\theta_L,\theta_L) + p_L \times BS_2(\theta_L,\theta_L) + V_1(NS;(\theta_L,\theta_L)) + V_2(NS;(\theta_L,\theta_L)).$$

  which lies below the welfare under the separating PBE, $V_1(NS;(\theta_L,\theta_L))+V_2(NS;(\theta_L,\theta_L))$, if

$$BS_1(\theta_L,\theta_L) > BS_2(\theta_L,\theta_L)$$

which does not hold since $BS_1(\theta_L, \theta_L) < 0$.

- Therefore, social welfare in the semiseparating PBE is lower than in the separating equilibrium, yielding a complete ranking of $SW_{separ} > SW_{semisepar} > SW_{pooling}$.

Complete information. Let us first compare social welfare when both countries' types are high. Under complete information, social welfare is $SW^{HH}_{complete} \equiv V_1(S, S; (\theta_H, \theta_H)) + V_2(S, S; (\theta_H, \theta_H))$, which coincides with equilibrium welfare under the pooling and separating PBE. Regarding the case where the leader's type is high but the follower's is low, $SW^{HL}_{complete} \equiv V_1(S, S; (\theta_H, \theta_L)) + V_2(S, S; (\theta_H, \theta_L))$ which also coincides with that under the pooling and separating equilibrium. When the leader's type is low but the follower's is high, $SW^{LH}_{complete} \equiv V_1(NS; (\theta_L, \theta_H)) + V_2(NS; (\theta_L, \theta_H))$, which coincides with the social welfare under the separating equilibrium, and hence, lies weakly below that under the pooling PBE if

$$V_1(NS; (\theta_L, \theta_H)) + V_2(NS; (\theta_L, \theta_H)) \leq V_1(S, S; (\theta_L, \theta_H)) + V_2(S, S; (\theta_L, \theta_H)).$$

Finally, when both countries' types are low, $SW^{LL}_{complete} \equiv V_1(NS; (\theta_L, \theta_L)) + V_2(NS; (\theta_L, \theta_L))$, which coincides with social welfare under the separating equilibrium, and therefore, lies above that under the pooling PBE. ■

7.7 *Proof of Proposition 5*

Separating equilibrium. Let us first show that the separating strategy profile in which the leader chooses to sign (not sign) the IEA when its type is high (low, respectively) cannot be supported as a PBE of the signaling game. Under such strategy profile, the m-th follower's beliefs are updated according to Bayes' rule and become $\mu_m(H|S) = 1$ and $\mu_m(H|NS) = 0$. Given these posterior beliefs, the m-th follower signs the agreement since $V_m(S, S_{n+1}, NS_{N-(n+1)}; \theta_H) > V_m(S, S_n, NS_{N-n}; \theta_H)$, since the followers' payoffs are increasing in the number of signatories. Given that this ranking of payoffs is valid for any given m-th follower and for any profile of n signatories and $N - n$ nonsignatories, it can be applied to all followers. Hence, all the N symmetric followers choose to sign the treaty. As a consequence, the leader chooses to participate in the treaty both when its type is high and low, since $V_1(S, S_N; \theta_K) > V_1(NS; \theta_K)$ for both θ_H and θ_L. But this strategy profile for the leader contradicts the separating strategy described above, and therefore it cannot be sustained as a PBE of the game.

Pooling equilibrium with signature. Let us next demonstrate that the pooling strategy profile in which the leader signs the IEA regardless of its type can be part of a PBE under certain conditions. In this strategy, the posterior beliefs of the last follower (in the N-th position) cannot be updated and thus coincide with its priors, i.e., $\mu_N(H|S) = p$ and $\mu_N(L|S) = 1 - p$. Given these beliefs, the N-th follower chooses to not sign the agreement, after observing that k_{N-1} countries participated in the treaty before the N-th position when it is called to move, if

$$\begin{aligned} & p \times V_N(S, S_{k_{N-1}+1}, NS_{(N-(k_{N-1}+1))}; \theta_H) + (1-p) \times V_N(S, S_{k_{N-1}+1}, NS_{(N-(k_{N-1}+1))}; \theta_L) \\ < \ & p \times V_N(S, S_{k_{N-1}}, NS_{(N-k_{N-1})}; \theta_H) + (1-p) \times V_N(S, S_{k_{N-1}}, NS_{(N-k_{N-1})}; \theta_L), \text{ or if} \end{aligned} \quad (1)$$

$$p < \frac{-BS_N((\theta_L, \theta_H), k_{N-1})}{BS_N((\theta_H, \theta_H), k_{N-1}) - BS_N((\theta_L, \theta_H), k_{N-1})} \equiv \widehat{p}(k_{N-1})$$

where, for all $K = \{H, L\}$, $BS_N((\theta_K, \theta_H), k_{N-1})$ denotes this follower's benefit from joining a treaty with k_{N-1} signatories, that is

$$BS_N((\theta_K, \theta_H), k_{N-1}) \equiv V_N(S, S_{k_{N-1}+1}, NS_{(N-(k_{N-1}+1))}; \theta_K) - V_N(S, S_{k_{N-1}}, NS_{(N-k_{N-1})}; \theta_K)$$

First, note that the above cutoff $\widehat{p}(k_{N-1})$ is positive, $\widehat{p}(k_{N-1}) > 0$, since the follower's payoffs satisfy $BS_N((\theta_L, \theta_H), k_{N-1}) < 0$ and $BS_N((\theta_H, \theta_H), k_{N-1}) > BS_N((\theta_L, \theta_H), k_{N-1})$ by definition. Second, $\widehat{p}(k_{N-1}) < 1$ given that

$$BS_N((\theta_H, \theta_H), k_{N-1}) - BS_N((\theta_L, \theta_H), k_{N-1}) > -BS_N((\theta_L, \theta_H), k_{N-1})$$

or $BS_N((\theta_H, \theta_H), k_{N-1}) > 0$, which holds for all k_{N-1}. Hence, when $p < \widehat{p}(k_{N-1})$, the last follower (in the N^{th} position) does not sign the treaty if it observes k_{N-1} signatories before his turn. Otherwise, he participates. Furthermore, note that since $BS_N((\theta_L, \theta_H), k_{N-1})$ is constant in the number of signatories but $BS_N((\theta_H, \theta_H), k_{N-1})$ is decreasing, cutoff $\widehat{p}(k_{N-1})$ becomes increasing in the number of signatories. As the following figure indicates, there must exist a number of signatories k^* for which $p = \widehat{p}(k^*)$, i.e., for all $k_{N-1} > k^*$ we have $p \leq \widehat{p}(k_{N-1})$ and therefore the N^{th}-follower does not participate, while for all $k_{N-1} < k^*$ we have that $p > \widehat{p}(k_{N-1})$ and the N^{th}-follower participates.

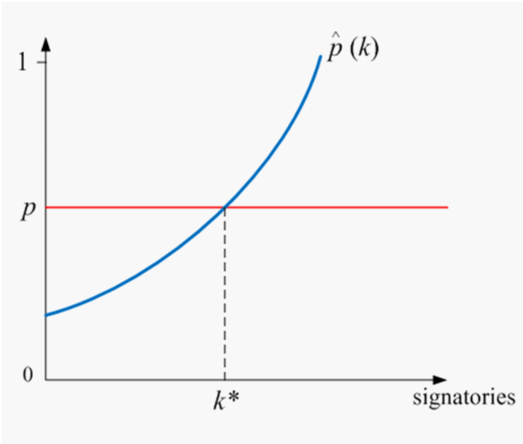

Figure A. Cutoff $\widehat{p}(k)$

$(N-1)$-*follower.* Let us now examine the $(N-1)$-follower. Its decision on whether to participate depends on the actions it anticipates for the N^{th}-follower afterwards.

1. If priors satisfy $p \leq \widehat{p}(k_{N-1})$, the $(N-1)$-follower anticipates that the N^{th}-follower will not sign the treaty afterwards. Therefore, the $(N-1)$-follower does not participate in the treaty, after observing that k_{N-2} countries participated in the treaty before its $(N-1)$-th position, if

$$\begin{aligned} & p \times V_{N-1}(S, S_{k_{N-2}+1}, NS_{(N-(k_{N-2}+1))}; \theta_H | NS_{k_N}) \\ & +(1-p) \times V_{N-1}(S, S_{k_{N-2}+1}, NS_{(N-(k_{N-2}+1))}; \theta_L | NS_{k_N}) \\ < \; & p \times V_{N-1}(S, S_{k_{N-2}}, NS_{(N-k_{N-2})}; \theta_H | NS_{k_N}) \\ & +(1-p) \times V_{N-1}(S, S_{k_{N-2}}, NS_{(N-k_{N-2})}; \theta_L | NS_{k_N}), \text{ or if} \end{aligned} \tag{2}$$

$$p < \frac{-BS_{N-1}((\theta_L, \theta_H), k_{N-2})}{BS_{N-1}((\theta_H, \theta_H), k_{N-2}) - BS_{N-1}((\theta_L, \theta_H), k_{N-2})} \equiv \widehat{p}(k_{N-2})$$

where the number of signatories at stage $N-1$, k_{N-2}, must satisfy $k_{N-2} > k^*$, where k^* solves $p = \widehat{p}(k^*)$. In this case, the $(N-1)$-follower's decision does not affect the N^{th}-follower's not signature afterwards, since $p \leq \widehat{p}(k_{N-1})$ for all $k_{N-1} > k_{N-2} > k^*$. In addition, note that since $BS_{N-1}((\theta_H, \theta_H), k_{N-2})$ decreases in the number of signatory countries, $\widehat{p}(k_{N-2}) \leq \widehat{p}(k_{N-1})$.

2. If, in contrast, priors satisfy $p > \widehat{p}(k_{N-1})$, the $(N-1)$-follower anticipates that the N^{th}-follower will sign the treaty afterwards. In this case the $(N-1)$-follower does not participate in the treaty if

$$\begin{aligned} & p \times V_{N-1}(S, S_{k_{N-2}+1}, NS_{(N-(k_{N-2}+1))}; \theta_H | S_{k_N}) \\ & +(1-p) \times V_{N-1}(S, S_{k_{N-2}+1}, NS_{(N-(k_{N-2}+1))}; \theta_L | S_{k_N}) \\ < \; & p \times V_{N-1}(S, S_{k_{N-2}}, NS_{(N-k_{N-2})}; \theta_H | S_{k_N}) \\ & +(1-p) \times V_{N-1}(S, S_{k_{N-2}}, NS_{(N-k_{N-2})}; \theta_L | S_{k_N}), \end{aligned} \tag{3}$$

where the payoffs in this expression coincide with those in (1) since they both describe the same number of signatory countries at the end of the game. We can therefore conclude that, in this case, the $(N-1)$-follower does not participate if $p \leq \widehat{p}(k_{N-1})$, as for the N^{th}-follower. Otherwise, if $p > \widehat{p}(k_{N-1})$, both the N^{th} and the $(N-1)$-follower participate. Finally, note that if k_{N-2} satisfies $k^* - 1 > k_{N-2} > k^*$, signing the treaty at stage $(N-1)$ induces the N^{th}-follower to not participate in the agreement in the next stage, while not signing it induces the N^{th}-follower to participate. In this case, the $(N-1)$-follower chooses not to sign since

$$\begin{aligned}
& p \times V_{N-1}(S, S_{k_{N-2}+1}, NS_{(N-(k_{N-2}+1))}; \theta_H | NS_{k_N}) \\
& +(1-p) \times V_{N-1}(S, S_{k_{N-2}+1}, NS_{(N-(k_{N-2}+1))}; \theta_L | NS_{k_N}) \\
< \; & p \times V_{N-1}(S, S_{k_{N-2}}, NS_{(N-k_{N-2})}; \theta_H | S_{k_N}) \\
& +(1-p) \times V_{N-1}(S, S_{k_{N-2}}, NS_{(N-k_{N-2})}; \theta_L | S_{k_N})
\end{aligned}$$

given that $V_{N-1}(S, S_{k_{N-2}+1}, NS_{(N-(k_{N-2}+1))}; \theta_K | NS_{k_N}) < V_{N-1}(S, S_{k_{N-2}}, NS_{(N-k_{N-2})}; \theta_K | S_{k_N})$ for all $K = \{H, L\}$ by definition.

$(N-2)$-follower. Let us continue with the $(N-2)$-follower. We must analyze three cases:

1. Both the $(N-1)$- and N^{th}-followers sign the treaty after country $N-2$. Note that this occurs when priors p are sufficiently high to satisfy $p > \widehat{p}(k_{N-1})$. Then the $(N-2)$-follower's expected utility comparison becomes similar to that in expression (3) above since it describes the same number of signatory countries at the end of the game. In this case, the $(N-2)$-follower chooses to not participate in the treaty according to the cutoff strategy $p \leq \widehat{p}(k_{N-1})$.

2. The $(N-1)$-follower signs after the $(N-2)$ country, but the N^{th}-follower does not. This occurs when, in particular, priors satisfy $p \in [\widehat{p}(k_{N-2}), \widehat{p}(k_{N-1})]$, where $\widehat{p}(k_{N-2}) \leq \widehat{p}(k_N - 1)$. In this case, the $(N-2)$-follower's expected utility comparison becomes similar to that in expression (2) above. Therefore, the $(N-2)$-follower chooses to not participate in the treaty according to the cutoff strategy $p \leq \widehat{p}(k_{N-2})$.

3. Neither the $(N-1)$- nor the N^{th}-followers sign the treaty after him. This occurs when, specifically, priors are low enough, i.e., $p < \widehat{p}(k_{N-2})$. In this case, the $(N-2)$-follower expected utility comparison is different from those examined above for the N^{th} and the $(N-1)$-follower, since two fewer followers are signing the agreement. Therefore, the $(N-2)$-follower chooses to not participate in the treaty according to the cutoff strategy $p \leq \widehat{p}(k_{N-3})$.

Importantly, the only case that becomes new relative to the N^{th} and $(N-1)$-followers (and where we obtain a cutoff strategy that was not used by these two followers) is the case in which no follower signs the agreement after the $(N-2)$-follower, finding cutoff strategy $p \leq \widehat{p}(k_{N-3})$. This procedure can be generalized to any follower choosing whether to participate in the treaty before the $(N-2)$-follower. We can extend a similar argument to all previous followers, whereby a given follower in position t participates in the treaty, after observing that k_{t-1} countries participated in the treaty, if

$$\begin{aligned}
& p \times V_t(S, S_{k_{t-1}+1}, NS_{(N-(k_{t-1}+1)}; \theta_H | NS_{(N-(k_{t-1}+1)}) \\
& +(1-p) \times V_t(S, S_{k_{t-1}+1}, NS_{(N-(k_{t-1}+1)}; \theta_L | NS_{(N-(k_{t-1}+1)}) \\
< \; & p \times V_t(S, S_{k_{t-1}}, NS_{(N-_{k_{t-1}})}; \theta_H | NS_{(N-(k_{t-1}+1))}) \\
& +(1-p) \times V_t(S, S_{k_{t-1}}, NS_{(N-_{k_{t-1}})}; \theta_L | NS_{(N-(k_{t-1}+1))})
\end{aligned}$$

We can hence conclude that, for a given prior p, all N followers sign the treaty if $p > \widehat{p}(k_{N-1})$, only the last follower chooses to not participate if $p \in [\widehat{p}(k_{N-2}), \widehat{p}(k_{N-1})]$, only the last two followers decide to not sign if $p \in [\widehat{p}(k_{N-3}), \widehat{p}(k_{N-2})]$, etc. Note that, as a consequence, the treaty is not signed by any follower if p is sufficiently low to satisfy $p \leq \widehat{p}(k_0)$.

If $p \leq \widehat{p}(k_0)$ then no follower signs the treaty, and the high-type leader participates while the low-type leader does not. This strategy profile hence cannot be supported as a pooling PBE of the game. If, in contrast, $p > \widehat{p}(k_0)$, one or more followers sign the treaty. Thus, both the high- and low-type leaders participate in the agreement. Therefore, this strategy profile can be sustained as a pooling PBE.

Pooling equilibrium with no signature. Finally, let us show that the pooling strategy profile where the leader does not sign the IEA regardless of its type cannot be sustained as part of a PBE. First, note that the follower's posterior beliefs cannot be updated using Bayes' rule, and hence must be arbitrarily specified, $\mu_J(H|S) = 1$ for any J-type follower. Given these beliefs, the follower responds signing the agreement. Hence, when the leader is a high type, it signs the agreement since $V_1(S, S_N; \theta_H) > V_1(NS; \theta_H)$. Therefore, the high-type leader signs the treaty and the pooling strategy profile in which no type of leader signs cannot be supported as PBE.

Intuitive Criterion. Let us apply the Cho and Kreps' (1987) Intuitive Criterion for the pooling PBE where $p > \widehat{p}(k_0)$, i.e., at least one follower signs the treaty. We first check if a deviation towards "not sign" is equilibrium dominated for either type of leader. When the leader is a high type, the highest payoff that it obtains by deviating towards "not sign" is $V_1(NS; \theta_H)$ which does not exceed its equilibrium payoff $V_1(S, S_n, NS_{N-n}; \theta_H)$, where $n \geq 1$. Similarly, the highest payoff that a low-type leader can obtain is $V_1(NS; \theta_L)$ which does not exceed its equilibrium payoff of $V_1(S, S_n, NS_{N-n}; \theta_L)$ since at least one follower signs the agreement. Hence, no type of leader has incentives to deviate towards "not sign." As a consequence, the follower's posterior beliefs cannot be updated after observing the off-the-equilibrium message "not sign", and hence the pooling PBE survives the Intuitive Criterion.

Special case: Consider the case in which followers' incentives to participate in the treaty are constant in the number of signatories, i.e., $BS_t((\theta_K, \theta_H), n)$ is constant in n. In this case cutoff $\widehat{p}(k_{t-1})$ becomes constant in the number of signatories that follower t observes before him, k_{t-1}. Let $\widehat{p}$ denote this cutoff. Hence, the N^{th}-follower uses the same cutoff strategy as the $(N-1)$-follower, and similarly for all previous followers. When $p \geq \widehat{p}$, the t-th follower signs the treaty, and similarly for all other symmetric followers. Provided that $V_1(S, S_N; \theta_K) > V_1(NS; \theta_K)$ for both types of leader, the leader participates in the agreement both when its type is high and low. Hence, the pooling strategy profile can be supported as PBE of the game when $p \geq \widehat{p}$. ■

7.8 *Proof of Proposition 6*

In the semiseparating equilibrium at least one follower must be mixing. Otherwise, the leader could anticipate all followers' actions and play pure strategies as in any of the strategy profiles described in proposition 5, which are not PBE of the signaling game when $p < \widehat{p}(k_0)$. Hence, at least one follower (in position t) must be indifferent between signing and not signing the treaty, that is,

$$\begin{aligned} & \mu_t(H|S) \times EV_t(S;\theta_H) + (1-\mu_t(H|S)) \times EV_t(S;\theta_L) \\ = \quad & \mu_t(H|S) \times EV_t(NS;\theta_H) + (1-\mu_t(H|S)) \times EV_t(NS;\theta_L), \end{aligned}$$

where $EV_t(S;\theta_K)$ denotes the expected equilibrium payoff that follower in position t obtains from participating in the agreement, after observing that k_{t-1} countries participated until stage $t-1$, when the true type of the leader is θ_K. In particular,

$$EV_t(S;\theta_K) \equiv \sum_{j=t+1}^{N} r_j(k_{j-1}) \times V_t(S, S_{k_{t-1}+1}, NS_{N-(k_{t-1}+1)}, S_j; \theta_K)$$

where $r_j(k_{j-1})$ denotes the probability that posterior followers, in position $j = \{t+1, ..., N\}$ choose to sign after observing k_{j-1} signatories before stage j. Similarly for the case of not signing, where $EV_t(NS;\theta_K) \equiv \sum_{j=t+1}^{N} r_j(k'_{j-1}) \times V_t(S, S_{k_{t-1}}, NS_{N-k_{t-1}}, S_j; \theta_K)$, where for completeness we allow the randomization of follower j after observing that the previous country (follower t) did not sign the treaty to differ from that in which follower t chooses to participate in the agreement. Hence, follower in position t is indifferent when

$$\mu_t(H|S) = \frac{EV_t(NS;\theta_L) - EV_t(S;\theta_L)}{[EV_t(S;\theta_H) - EV_t(NS;\theta_H)] + [EV_t(NS;\theta_L) - EV_t(S;\theta_L)]} \equiv \widetilde{p}_t$$

where $\widetilde{p}_t > 0$ since $EV_t(NS;\theta_L) > EV_t(S;\theta_L)$ by definition when follower t knows that the leader is a low-type. In addition, $\widetilde{p}_t < 1$ given that $EV_t(S;\theta_H) > EV_t(NS;\theta_H)$ when follower t knows that the leader's type is high. We can next use the follower's posterior beliefs in order to find the probability with which the leader randomizes when its type is low, p_L, by using Bayes' rule.

$$\mu_t(H|S) = \widehat{p}_t = \frac{(1-p) \times p_L}{((1-p) \times p_L) + (p \times p_H)}$$

where $p_H = 1$. Solving for p_L, we obtain $p_L = \frac{p}{1-p}\frac{\widetilde{p}_t}{1-\widetilde{p}_t}$, which is positive, $p_L > 0$, since $p, \widetilde{p}_t \in (0,1)$. In addition, $p_L < 1$ for all probability p satisfying $p < 1 - \widetilde{p}_t$. There exists a number of followers $\overline{N}$ for which $|BS_t((\theta_H, \theta_H), n)| > |BS_t((\theta_L, \theta_H), n)|$ for any number of signatories $n \leq \overline{N}$. For any follower $t \leq \overline{N}$, cutoff $\widetilde{p}_t$ satisfies $\widetilde{p}_t < 1/2$, and then $p < 1 - \widetilde{p}_t$ holds for all priors p since $p < \widetilde{p}_t$ by assumption. If, in contrast, the number of followers is relatively high, and $|BS_t((\theta_H, \theta_H), n)| < |BS_t((\theta_L, \theta_H), n)|$ holds for some n, then the semiseparating equilibrium can be sustained if $p < 1 - \widetilde{p}_t$ holds for all follower t.

Finally, note that if the leader mixes with probability $p_L \in (0,1)$, it must be that followers make the leader indifferent between signing and not signing the agreement. That is, there is a vector of followers' mixed strategies $r = (r_1, r_2, ..., r_N)$ that solves

$$\sum_{j=1}^{N} r_j \times V_1(S, S_j; \theta_L) = V_1(NS; \theta_L)$$

where $\sum_{j=1}^{N} r_j \times V_1(S, S_j; \theta_L)$ represents the leader's expected equilibrium payoff from participating in the treaty, where N followers randomize their participation afterwards. ■

References

[1] Bagwell, K. (2009) "Self-Enforcing Trade Agreements and Private Information," NBER Working Paper, 14812.

[2] Barrett, S. (1992) "Convention on Climate Change - Economic Aspects of Negotiations," Paris, OECD.

[3] Barrett, S. (1994a) "Strategic Environmental Policy and International Trade," Journal of Public Economics, 54, pp. 325–38.

[4] Barrett, S. (1994b) "Self-enforcing international environmental agreements," Oxford Economic Papers 46, pp. 878–94.

[5] Barrett, S. (1999) "A theory of full international cooperation," Journal of Theoretical Politics 11, pp. 519–41.

[6] Carraro, C., and D. Siniscalco. (2001) Transfers, Commitments and Issue Linkage in International Environmental Negotiations, in A. Ulph, ed., *Environmental Policy, International Agreements and International Trade*, Oxford University Press.

[7] Cesar, H. (1994) *Control and game models of the greenhouse effect*, Berlin: Springer-Verlag.

[8] Cho, I. and D. Kreps (1987) "Signaling games and stable equilibrium," Quarterly Journal of Economics, 102, 179-222.

[9] Downs, G., D. Rocke, and P. Barsoom (1996) "Is the good news about compliance good news about cooperation?," International Organization, 50(3), 379–406.

[10] Ederington, J. (2002) "Trade and Domestic Policy Linkage in International Agreements," International Economic Review, 43(4), pp. 1347–68.

[11] Iida, K. (1993) "When and How Do Domestic Constraints Matter?: Two-Level Games with Uncertainty," Journal of Conflict Resolution, 37, pp. 403-26.

[12] Kreps, D., P. Milgrom, J. Roberts and R. Wilson (1982) "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma," Journal of Economic Theory 27, pp. 245-52.

[13] Lee, G. M. (2007), "Trade Agreements with Domestic Policies As Disguised Protection," Journal of International Economics, 71.1, 241-59.

[14] Martin, A. and W. Vergote (2008), "On the Role of Retaliation in Trade Agreements," Journal of International Economics, 76.1, 61-77.

[15] Von Stein, J. (2008). "The International Law and Politics of Climate Change: Ratification of the United Nations Framework Convention and the Kyoto Protocol," Journal of Conflict Resolution, 52(2), pp. 243-68.

[16] Whalley, J. (1991) "The Interface Between Environmental and Trade Policies," The Economic Journal, 101, pp. 180-89.